Elementary Linear Algebra

Kuttler

January 23, 2009

Contents

Introduction

This is an introduction to linear algebra. The main part of the book features row operations and everything is done in terms of the row reduced echelon form and specific algorithms. At the end, the more abstract notions of vector spaces and linear transformations on vector spaces are presented. However, this is intended to be a first course in linear algebra for students who are sophomores or juniors who have had a course in one variable calculus and a reasonable background in college algebra. I have given complete proofs of all the fundamental ideas but some topics such as Markov matrices are not complete in this book but receive a plausible introduction. The book contains a complete treatment of determinants and a simple proof of the Cayley Hamilton theorem although these are optional topics. The Jordan form is presented as an appendix. I see this theorem as the beginning of more advanced topics in linear algebra and not really part of a beginning linear algebra course. There are extensions of many of the topics of this book in my on line book [9]. I have also not emphasized that linear algebra can be carried out with any field although I have done everything in terms of either the real numbers or the complex numbers. It seems to me this is a reasonable specialization for a first course in linear algebra.

$\mathbb{F}^n$

2.0.1 Outcomes

A. Understand the symbol, $\mathbb{F}^n$ in the case where $\mathbb{F}$ equals the real numbers, $\mathbb{R}$ or the complex numbers, $\mathbb{C}$.

B. Know how to do algebra with vectors in $\mathbb{F}^n$, including vector addition and scalar multiplication.

C. Understand the geometric significance of an element of $\mathbb{F}^n$ when possible.

The notation, $\mathbb{C}^n$ refers to the collection of ordered lists of n complex numbers. Since every real number is also a complex number, this simply generalizes the usual notion of $\mathbb{R}^n$, the collection of all ordered lists of n real numbers. In order to avoid worrying about whether it is real or complex numbers which are being referred to, the symbol $\mathbb{F}$ will be used. If it is not clear, always pick $\mathbb{C}$.

Definition 2.0.1 *Define* $\mathbb{F}^n \equiv \{(x_1, \cdots, x_n) : x_j \in \mathbb{F} \text{ for } j = 1, \cdots, n\}$.

$$(x_1, \cdots, x_n) = (y_1, \cdots, y_n)$$

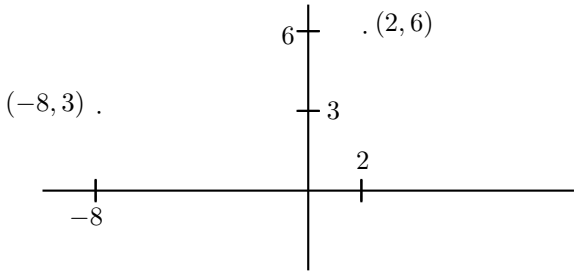
if and only if for all $j = 1, \cdots, n$, $x_j = y_j$. *When* $(x_1, \cdots, x_n) \in \mathbb{F}^n$, *it is conventional to denote* $(x_1, \cdots, x_n)$ *by the single bold face letter,* $\mathbf{x}$. *The numbers,* x_j *are called the coordinates. The set*

$$\{(0, \cdots, 0, t, 0, \cdots, 0) : t \in \mathbb{F}\}$$

for t *in the* i^{th} *slot is called the* i^{th} *coordinate axis. The point* $\mathbf{0} \equiv (0, \cdots, 0)$ *is called the origin. Elements in* $\mathbb{F}^n$ *are called* ***vectors.***

Thus $(1, 2, 4i) \in \mathbb{F}^3$ and $(2, 1, 4i) \in \mathbb{F}^3$ but $(1, 2, 4i) \neq (2, 1, 4i)$ because, even though the same numbers are involved, they don't match up. In particular, the first entries are not equal.

The geometric significance of $\mathbb{R}^n$ for $n \leq 3$ has been encountered already in calculus or in pre-calculus. Here is a short review. First consider the case when $n = 1$. Then from the definition, $\mathbb{R}^1 = \mathbb{R}$. Recall that $\mathbb{R}$ is identified with the points of a line. Look at the number line again. Observe that this amounts to identifying a point on this line with a real number. In other words a real number determines where you are on this line. Now suppose $n = 2$ and consider two lines which intersect each other at right angles as shown in the following picture.



Notice how you can identify a point shown in the plane with the ordered pair, $(2,6)$. You go to the right a distance of 2 and then up a distance of 6. Similarly, you can identify another point in the plane with the ordered pair $(-8,3)$. Go to the left a distance of 8 and then up a distance of 3. The reason you go to the left is that there is a $-$ sign on the eight. From this reasoning, every ordered pair determines a unique point in the plane. Conversely, taking a point in the plane, you could draw two lines through the point, one vertical and the other horizontal and determine unique points, x_1 on the horizontal line in the above picture and x_2 on the vertical line in the above picture, such that the point of interest is identified with the ordered pair, (x_1, x_2). In short, points in the plane can be identified with ordered pairs similar to the way that points on the real line are identified with real numbers. Now suppose $n = 3$. As just explained, the first two coordinates determine a point in a plane. Letting the third component determine how far up or down you go, depending on whether this number is positive or negative, this determines a point in space. Thus, $(1, 4, -5)$ would mean to determine the point in the plane that goes with $(1, 4)$ and then to go below this plane a distance of 5 to obtain a unique point in space. You see that the ordered triples correspond to points in space just as the ordered pairs correspond to points in a plane and single real numbers correspond to points on a line.

You can't stop here and say that you are only interested in $n \leq 3$. What if you were interested in the motion of two objects? You would need three coordinates to describe where the first object is and you would need another three coordinates to describe where the other object is located. Therefore, you would need to be considering $\mathbb{R}^6$. If the two objects moved around, you would need a time coordinate as well. As another example, consider a hot object which is cooling and suppose you want the temperature of this object. How many coordinates would be needed? You would need one for the temperature, three for the position of the point in the object and one more for the time. Thus you would need to be considering $\mathbb{R}^5$. Many other examples can be given. Sometimes n is very large. This is often the case in applications to business when they are trying to maximize profit subject to constraints. It also occurs in numerical analysis when people try to solve hard problems on a computer.

There are other ways to identify points in space with three numbers but the one presented is the most basic. In this case, the coordinates are known as Cartesian coordinates after Descartes[1] who invented this idea in the first half of the seventeenth century. I will often not bother to draw a distinction between the point in space and its Cartesian coordinates.

The geometric significance of $\mathbb{C}^n$ for $n > 1$ is not available because each copy of $\mathbb{C}$ corresponds to the plane or $\mathbb{R}^2$.

[1]René Descartes 1596-1650 is often credited with inventing analytic geometry although it seems the ideas were actually known much earlier. He was interested in many different subjects, physiology, chemistry, and physics being some of them. He also wrote a large book in which he tried to explain the book of Genesis scientifically. Descartes ended up dying in Sweden.

2.1 Algebra in $\mathbb{F}^n$

There are two algebraic operations done with elements of $\mathbb{F}^n$. One is addition and the other is multiplication by numbers, called scalars. In the case of $\mathbb{C}^n$ the scalars are complex numbers while in the case of $\mathbb{R}^n$ the only allowed scalars are real numbers. Thus, the scalars always come from $\mathbb{F}$ in either case.

Definition 2.1.1 *If $\mathbf{x} \in \mathbb{F}^n$ and $a \in \mathbb{F}$, also called a scalar, then $a\mathbf{x} \in \mathbb{F}^n$ is defined by*

$$a\mathbf{x} = a(x_1, \cdots, x_n) \equiv (ax_1, \cdots, ax_n). \tag{2.1}$$

This is known as scalar multiplication. If $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$ then $\mathbf{x} + \mathbf{y} \in \mathbb{F}^n$ and is defined by

$$\begin{aligned} \mathbf{x} + \mathbf{y} &= (x_1, \cdots, x_n) + (y_1, \cdots, y_n) \\ &\equiv (x_1 + y_1, \cdots, x_n + y_n) \end{aligned} \tag{2.2}$$

$\mathbb{F}^n$ is often called n dimensional space. With this definition, the algebraic properties satisfy the conclusions of the following theorem.

Theorem 2.1.2 *For $\mathbf{v}, \mathbf{w} \in \mathbb{F}^n$ and α, β scalars, (real numbers), the following hold.*

$$\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}, \tag{2.3}$$

the commutative law of addition,

$$(\mathbf{v} + \mathbf{w}) + \mathbf{z} = \mathbf{v} + (\mathbf{w} + \mathbf{z}), \tag{2.4}$$

the associative law for addition,

$$\mathbf{v} + \mathbf{0} = \mathbf{v}, \tag{2.5}$$

the existence of an additive identity,

$$\mathbf{v} + (-\mathbf{v}) = \mathbf{0}, \tag{2.6}$$

the existence of an additive inverse, Also

$$\alpha(\mathbf{v} + \mathbf{w}) = \alpha\mathbf{v} + \alpha\mathbf{w}, \tag{2.7}$$

$$(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}, \tag{2.8}$$

$$\alpha(\beta\mathbf{v}) = \alpha\beta(\mathbf{v}), \tag{2.9}$$

$$1\mathbf{v} = \mathbf{v}. \tag{2.10}$$

In the above $\mathbf{0} = (0, \cdots, 0)$.

You should verify these properties all hold. For example, consider 2.7

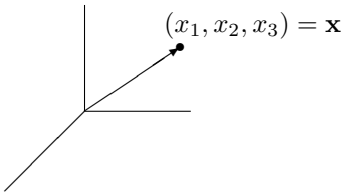
$$\begin{aligned} \alpha(\mathbf{v} + \mathbf{w}) &= \alpha(v_1 + w_1, \cdots, v_n + w_n) \\ &= (\alpha(v_1 + w_1), \cdots, \alpha(v_n + w_n)) \\ &= (\alpha v_1 + \alpha w_1, \cdots, \alpha v_n + \alpha w_n) \\ &= (\alpha v_1, \cdots, \alpha v_n) + (\alpha w_1, \cdots, \alpha w_n) \\ &= \alpha\mathbf{v} + \alpha\mathbf{w}. \end{aligned}$$

As usual subtraction is defined as $\mathbf{x} - \mathbf{y} \equiv \mathbf{x} + (-\mathbf{y})$.

2.2 Geometric Meaning Of Vectors

The geometric meaning is especially significant in the case of $\mathbb{R}^n$ for $n = 2, 3$. Here is a short discussion of this topic.

Definition 2.2.1 *Let* $\mathbf{x} = (x_1, \cdots, x_n)$ *be the coordinates of a point in* $\mathbb{R}^n$. *Imagine an arrow with its tail at* $\mathbf{0} = (0, \cdots, 0)$ *and its point at* $\mathbf{x}$ *as shown in the following picture in the case of* $\mathbb{R}^3$.

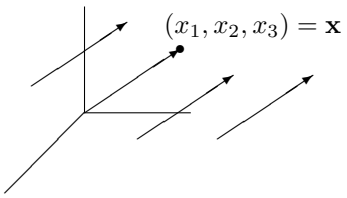


Then this arrow is called the ***position vector*** *of the point,* $\mathbf{x}$. *Given two points,* P, Q *whose coordinates are* $(p_1, \cdots, p_n)$ *and* $(q_1, \cdots, q_n)$ *respectively, one can also determine the position vector from* P *to* Q *defined as follows.*

$$\overrightarrow{PQ} \equiv (q_1 - p_1, \cdots, q_n - p_n)$$

Thus every point determines a vector and conversely, every such vector (arrow) which has its tail at $\mathbf{0}$ determines a point of $\mathbb{R}^n$, namely the point of $\mathbb{R}^n$ which coincides with the point of the vector. Also two different points determine a position vector going from one to the other as just explained.

Imagine taking the above position vector and moving it around, always keeping it pointing in the same direction as shown in the following picture.



After moving it around, it is regarded as the same vector because it points in the same direction and has the same length.[2]Thus each of the arrows in the above picture is regarded as the same vector. The **components** of this vector are the numbers, $x_1, \cdots, x_n$. You should think of these numbers as directions for obtainng an arrow. Starting at some point, $(a_1, a_2, \cdots, a_n)$ in $\mathbb{R}^n$, you move to the point $(a_1 + x_1, \cdots, a_n)$ and from there to the point $(a_1 + x_1, a_2 + x_2, a_3 \cdots, a_n)$ and then to $(a_1 + x_1, a_2 + x_2, a_3 + x_3, \cdots, a_n)$ and continue this way until you obtain the point $(a_1 + x_1, a_2 + x_2, \cdots, a_n + x_n)$. The arrow having its tail at $(a_1, a_2, \cdots, a_n)$ and its point at $(a_1 + x_1, a_2 + x_2, \cdots, a_n + x_n)$ looks just like the arrow which has its tail at $\mathbf{0}$ and its point at $(x_1, \cdots, x_n)$ so it is regarded as the same vector.

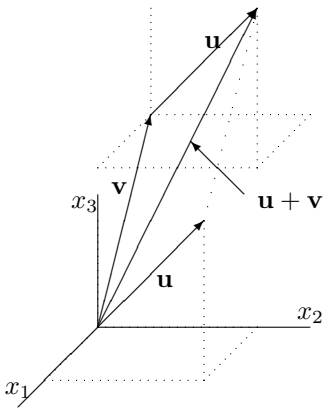
2.3 Geometric Meaning Of Vector Addition

It was explained earlier that an element of $\mathbb{R}^n$ is an n tuple of numbers and it was also shown that this can be used to determine a point in three dimensional space in the case

[2]I will discuss how to define length later. For now, it is only necessary to observe that the length should be defined in such a way that it does not change when such motion takes place.

where $n = 3$ and in two dimensional space, in the case where $n = 2$. This point was specified relative to some coordinate axes.

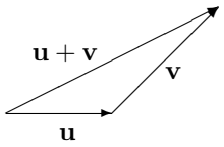
Consider the case where $n = 3$ for now. If you draw an arrow from the point in three dimensional space determined by $(0, 0, 0)$ to the point (a, b, c) with its tail sitting at the point $(0, 0, 0)$ and its point at the point (a, b, c), this arrow is called the **position vector** of the point determined by $\mathbf{u} \equiv (a, b, c)$. One way to get to this point is to start at $(0, 0, 0)$ and move in the direction of the x_1 axis to $(a, 0, 0)$ and then in the direction of the x_2 axis to $(a, b, 0)$ and finally in the direction of the x_3 axis to (a, b, c). It is evident that the same arrow (vector) would result if you began at the point, $\mathbf{v} \equiv (d, e, f)$, moved in the direction of the x_1 axis to $(d + a, e, f)$, then in the direction of the x_2 axis to $(d + a, e + b, f)$, and finally in the x_3 direction to $(d + a, e + b, f + c)$ only this time, the arrow would have its tail sitting at the point determined by $\mathbf{v} \equiv (d, e, f)$ and its point at $(d + a, e + b, f + c)$. It is said to be the same arrow (vector) because it will point in the same direction and have the same length. It is like you took an actual arrow, the sort of thing you shoot with a bow, and moved it from one location to another keeping it pointing the same direction. This is illustrated in the following picture in which $\mathbf{v} + \mathbf{u}$ is illustrated. Note the parallelogram determined in the picture by the vectors $\mathbf{u}$ and $\mathbf{v}$.



Thus the geometric significance of $(d, e, f) + (a, b, c) = (d + a, e + b, f + c)$ is this. You start with the position vector of the point (d, e, f) and at its point, you place the vector determined by (a, b, c) with its tail at (d, e, f). Then the point of this last vector will be $(d + a, e + b, f + c)$. This is the geometric significance of vector addition. Also, as shown in the picture, $\mathbf{u} + \mathbf{v}$ is the directed diagonal of the parallelogram determined by the two vectors $\mathbf{u}$ and $\mathbf{v}$. A similar interpretation holds in $\mathbb{R}^n, n > 3$ but I can't draw a picture in this case.

Since the convention is that identical arrows pointing in the same direction represent the same vector, the geometric significance of vector addition is as follows in any number of dimensions.

Procedure 2.3.1 *Let* $\mathbf{u}$ *and* $\mathbf{v}$ *be two vectors. Slide* $\mathbf{v}$ *so that the tail of* $\mathbf{v}$ *is on the point of* $\mathbf{u}$. *Then draw the arrow which goes from the tail of* $\mathbf{u}$ *to the point of the slid vector,* $\mathbf{v}$. *This arrow represents the vector* $\mathbf{u} + \mathbf{v}$.



Note that $P+\overrightarrow{PQ} = Q$.

2.4 Distance Between Points In $\mathbb{R}^n$ Length Of A Vector

How is distance between two points in $\mathbb{R}^n$ defined?

Definition 2.4.1 *Let* $\mathbf{x} = (x_1, \cdots, x_n)$ *and* $\mathbf{y} = (y_1, \cdots, y_n)$ *be two points in* $\mathbb{R}^n$. *Then* $|\mathbf{x} - \mathbf{y}|$ *to indicates the distance between these points and is defined as*

$$\text{distance between } \mathbf{x} \text{ and } \mathbf{y} \equiv |\mathbf{x} - \mathbf{y}| \equiv \left(\sum_{k=1}^{n} |x_k - y_k|^2 \right)^{1/2}.$$

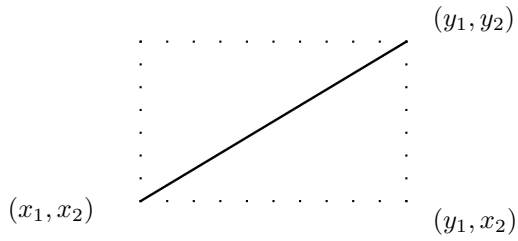
This is called the ***distance formula****. Thus* $|\mathbf{x}| \equiv |\mathbf{x} - \mathbf{0}|$. *The symbol,* $B(\mathbf{a}, r)$ *is defined by*

$$B(\mathbf{a}, r) \equiv \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x} - \mathbf{a}| < r\}.$$

This is called an ***open ball*** *of radius* r *centered at* $\mathbf{a}$. *It means all points in* $\mathbb{R}^n$ *which are closer to* $\mathbf{a}$ *than* r. *The length of a vector* $\mathbf{x}$ *is the distance between* $\mathbf{x}$ *and* $\mathbf{0}$.

First of all note this is a generalization of the notion of distance in $\mathbb{R}$. There the distance between two points, x and y was given by the absolute value of their difference. Thus $|x - y|$ is equal to the distance between these two points on $\mathbb{R}$. Now $|x - y| = \left((x - y)^2\right)^{1/2}$ where the square root is always the positive square root. Thus it is the same formula as the above definition except there is only one term in the sum. Geometrically, this is the right way to define distance which is seen from the Pythagorean theorem. Often people use two lines to denote this distance, $||\mathbf{x} - \mathbf{y}||$. However, I want to emphasize this is really just like the absolute value. Also, the notation I am using is fairly standard.

Consider the following picture in the case that $n = 2$.



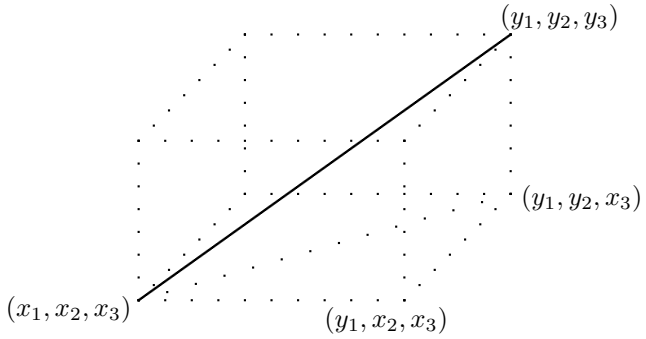
There are two points in the plane whose Cartesian coordinates are (x_1, x_2) and (y_1, y_2) respectively. Then the solid line joining these two points is the hypotenuse of a right triangle

which is half of the rectangle shown in dotted lines. What is its length? Note the lengths of the sides of this triangle are $|y_1 - x_1|$ and $|y_2 - x_2|$. Therefore, the Pythagorean theorem implies the length of the hypotenuse equals

$$\left(|y_1 - x_1|^2 + |y_2 - x_2|^2\right)^{1/2} = \left((y_1 - x_1)^2 + (y_2 - x_2)^2\right)^{1/2}$$

which is just the formula for the distance given above. In other words, this distance defined above is the same as the distance of plane geometry in which the Pythagorean theorem holds.

Now suppose $n = 3$ and let (x_1, x_2, x_3) and (y_1, y_2, y_3) be two points in $\mathbb{R}^3$. Consider the following picture in which one of the solid lines joins the two points and a dotted line joins the points (x_1, x_2, x_3) and (y_1, y_2, x_3).



By the Pythagorean theorem, the length of the dotted line joining (x_1, x_2, x_3) and (y_1, y_2, x_3) equals

$$\left((y_1 - x_1)^2 + (y_2 - x_2)^2\right)^{1/2}$$

while the length of the line joining (y_1, y_2, x_3) to (y_1, y_2, y_3) is just $|y_3 - x_3|$. Therefore, by the Pythagorean theorem again, the length of the line joining the points (x_1, x_2, x_3) and (y_1, y_2, y_3) equals

$$\begin{aligned}&\left\{\left[\left((y_1 - x_1)^2 + (y_2 - x_2)^2\right)^{1/2}\right]^2 + (y_3 - x_3)^2\right\}^{1/2}\\ &= \left((y_1 - x_1)^2 + (y_2 - x_2)^2 + (y_3 - x_3)^2\right)^{1/2},\end{aligned}$$

which is again just the distance formula above.

This completes the argument that the above definition is reasonable. Of course you cannot continue drawing pictures in ever higher dimensions but there is no problem with the formula for distance in any number of dimensions. Here is an example.

Example 2.4.2 *Find the distance between the points in* $\mathbb{R}^4$, $\mathbf{a} = (1, 2, -4, 6)$ *and* $\mathbf{b} = (2, 3, -1, 0)$

Use the distance formula and write

$$|\mathbf{a} - \mathbf{b}|^2 = (1 - 2)^2 + (2 - 3)^2 + (-4 - (-1))^2 + (6 - 0)^2 = 47$$

Therefore, $|\mathbf{a}-\mathbf{b}| = \sqrt{47}$.

All this amounts to defining the distance between two points as the length of a straight line joining these two points. However, there is nothing sacred about using straight lines. One could define the distance to be the length of some other sort of line joining these points. It won't be done in this book but sometimes this sort of thing is done.

Another convention which is usually followed, especially in $\mathbb{R}^2$ and $\mathbb{R}^3$ is to denote the first component of a point in $\mathbb{R}^2$ by x and the second component by y. In $\mathbb{R}^3$ it is customary to denote the first and second components as just described while the third component is called z.

Example 2.4.3 *Describe the points which are at the same distance between* $(1,2,3)$ *and* $(0,1,2)$.

Let (x,y,z) be such a point. Then

$$\sqrt{(x-1)^2+(y-2)^2+(z-3)^2} = \sqrt{x^2+(y-1)^2+(z-2)^2}.$$

Squaring both sides

$$(x-1)^2+(y-2)^2+(z-3)^2 = x^2+(y-1)^2+(z-2)^2$$

and so

$$x^2-2x+14+y^2-4y+z^2-6z = x^2+y^2-2y+5+z^2-4z$$

which implies

$$-2x+14-4y-6z = -2y+5-4z$$

and so

$$2x+2y+2z = -9. \tag{2.11}$$

Since these steps are reversible, the set of points which is at the same distance from the two given points consists of the points, (x,y,z) such that 2.11 holds.

There are certain properties of the distance which are obvious. Two of them which follow directly from the definition are

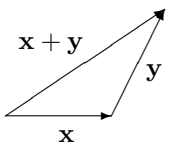
$$|\mathbf{x}-\mathbf{y}| = |\mathbf{y}-\mathbf{x}|,$$

$$|\mathbf{x}-\mathbf{y}| \geq 0 \text{ and equals 0 only if } \mathbf{y}=\mathbf{x}.$$

The third fundamental property of distance is known as the triangle inequality. Recall that in any triangle the sum of the lengths of two sides is always at least as large as the third side. I will show you a proof of this later. This is usually stated as

$$|\mathbf{x}+\mathbf{y}| \leq |\mathbf{x}|+|\mathbf{y}|.$$

Here is a picture which illustrates the statement of this inequality in terms of geometry.



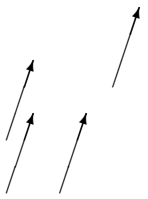
2.5 Geometric Meaning Of Scalar Multiplication

As discussed earlier, $\mathbf{x} = (x_1, x_2, x_3)$ determines a vector. You draw the line from $\mathbf{0}$ to $\mathbf{x}$ placing the point of the vector on $\mathbf{x}$. What is the length of this vector? The length of this vector is defined to equal $|\mathbf{x}|$ as in Definition 2.4.1. Thus the length of $\mathbf{x}$ equals $\sqrt{x_1^2 + x_2^2 + x_3^2}$. When you multiply $\mathbf{x}$ by a scalar, α, you get $(\alpha x_1, \alpha x_2, \alpha x_3)$ and the length of this vector is defined as $\sqrt{\left((\alpha x_1)^2 + (\alpha x_2)^2 + (\alpha x_3)^2\right)} = |\alpha| \sqrt{x_1^2 + x_2^2 + x_3^2}$. Thus the following holds.

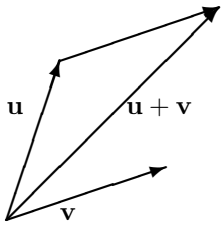
$$|\alpha \mathbf{x}| = |\alpha| \, |\mathbf{x}| .$$

In other words, multiplication by a scalar magnifies the length of the vector. What about the direction? You should convince yourself by drawing a picture that if α is negative, it causes the resulting vector to point in the opposite direction while if $\alpha > 0$ it preserves the direction the vector points.

You can think of vectors as quantities which have direction and magnitude, little arrows. Thus any two little arrows which have the same length and point in the same direction are considered to be the same vector even if their tails are at different points.

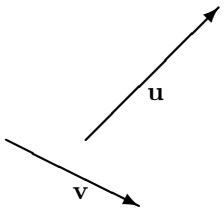


You can always slide such an arrow and place its tail at the origin. If the resulting point of the vector is (a, b, c), it is clear the length of the little arrow is $\sqrt{a^2 + b^2 + c^2}$. Geometrically, the way you add two geometric vectors is to place the tail of one on the point of the other and then to form the vector which results by starting with the tail of the first and ending with this point as illustrated in the following picture. Also when (a, b, c) is referred to as a vector, you mean any of the arrows which have the same direction and magnitude as the position vector of this point. Geometrically, for $\mathbf{u} = (u_1, u_2, u_3)$, $\alpha \mathbf{u}$ is any of the little arrows which have the same direction and magnitude as $(\alpha u_1, \alpha u_2, \alpha u_3)$.



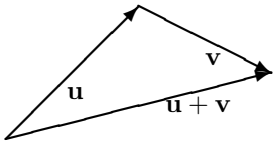
The following example is art which illustrates these definitions and conventions.

Exercise 2.5.1 *Here is a picture of two vectors,* $\mathbf{u}$ *and* $\mathbf{v}$.

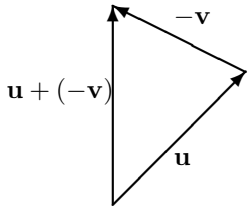


Sketch a picture of $\mathbf{u}+\mathbf{v}, \mathbf{u}-\mathbf{v}$, *and* $\mathbf{u}+2\mathbf{v}$.

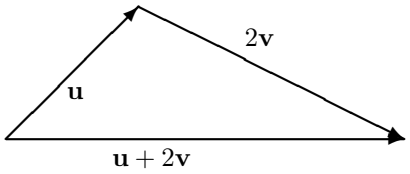
First here is a picture of $\mathbf{u}+\mathbf{v}$. You first draw $\mathbf{u}$ and then at the point of $\mathbf{u}$ you place the tail of $\mathbf{v}$ as shown. Then $\mathbf{u}+\mathbf{v}$ is the vector which results which is drawn in the following pretty picture.



Next consider $\mathbf{u}-\mathbf{v}$. This means $\mathbf{u}+(-\mathbf{v})$. From the above geometric description of vector addition, $-\mathbf{v}$ is the vector which has the same length but which points in the opposite direction to $\mathbf{v}$. Here is a picture.



Finally consider the vector $\mathbf{u}+2\mathbf{v}$. Here is a picture of this one also.

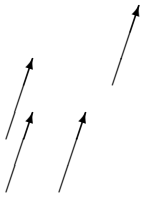


2.6 Vectors And Physics

Suppose you push on something. What is important? There are really two things which are important, how hard you push and the direction you push. This illustrates the concept of force.

Definition 2.6.1 ***Force*** *is a vector. The magnitude of this vector is a measure of how hard it is pushing. It is measured in units such as Newtons or pounds or tons. Its direction is the direction in which the push is taking place.*

Vectors are used to model force and other physical vectors like velocity. What was just described would be called a force vector. It has two essential ingredients, its magnitude and its direction. Geometrically think of vectors as directed line segments or arrows as shown in the following picture in which all the directed line segments are considered to be the same vector because they have the same direction, the direction in which the arrows point, and the same magnitude (length).



Because of this fact that only direction and magnitude are important, it is always possible to put a vector in a certain particularly simple form. Let $\overrightarrow{\mathbf{pq}}$ be a directed line segment or vector. Then it follows that $\overrightarrow{\mathbf{pq}}$ consists of the points of the form

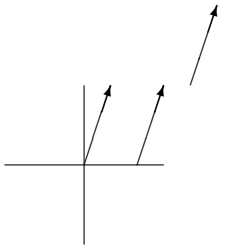
$$\mathbf{p} + t(\mathbf{q} - \mathbf{p})$$

where $t \in [0,1]$. Subtract $\mathbf{p}$ from all these points to obtain the directed line segment consisting of the points

$$\mathbf{0} + t(\mathbf{q} - \mathbf{p}),\ t \in [0,1].$$

The point in $\mathbb{R}^n, \mathbf{q} - \mathbf{p}$, will represent the vector.

Geometrically, the arrow, $\overrightarrow{\mathbf{pq}}$, was slid so it points in the same direction and the base is at the origin, $\mathbf{0}$. For example, see the following picture.



In this way vectors can be identified with points of $\mathbb{R}^n$.

Definition 2.6.2 *Let $\mathbf{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n$. The* ***position vector*** *of this point is the vector whose point is at $\mathbf{x}$ and whose tail is at the origin, $(0, \cdots, 0)$. If $\mathbf{x} = (x_1, \cdots, x_n)$ is called a vector, the vector which is meant is this position vector just described. Another term associated with this is* ***standard position.*** *A vector is in standard position if the tail is placed at the origin.*

It is customary to identify the point in $\mathbb{R}^n$ with its position vector.

The magnitude of a vector determined by a directed line segment $\overrightarrow{\mathbf{pq}}$ is just the distance between the point $\mathbf{p}$ and the point $\mathbf{q}$. By the distance formula this equals

$$\left(\sum_{k=1}^{n}\left(q_k - p_k\right)^2\right)^{1/2} = |\mathbf{p}-\mathbf{q}|$$

and for $\mathbf{v}$ any vector in $\mathbb{R}^n$ the magnitude of $\mathbf{v}$ equals $\left(\sum_{k=1}^{n} v_k^2\right)^{1/2} = |\mathbf{v}|$.

Example 2.6.3 *Consider the vector,* $\mathbf{v} \equiv (1,2,3)$ *in* $\mathbb{R}^n$. *Find* $|\mathbf{v}|$.

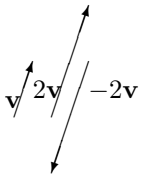
First, the vector is the directed line segment (arrow) which has its base at $\mathbf{0} \equiv (0,0,0)$ and its point at $(1,2,3)$. Therefore,

$$|\mathbf{v}| = \sqrt{1^2+2^2+3^2} = \sqrt{14}.$$

What is the geometric significance of scalar multiplication? If $\mathbf{a}$ represents the vector, $\mathbf{v}$ in the sense that when it is slid to place its tail at the origin, the element of $\mathbb{R}^n$ at its point is $\mathbf{a}$, what is $r\mathbf{v}$?

$$\begin{aligned}|r\mathbf{v}| &= \left(\sum_{k=1}^{n}\left(ra_i\right)^2\right)^{1/2} = \left(\sum_{k=1}^{n} r^2\left(a_i\right)^2\right)^{1/2} \\ &= \left(r^2\right)^{1/2}\left(\sum_{k=1}^{n} a_i^2\right)^{1/2} = |r|\,|\mathbf{v}|.\end{aligned}$$

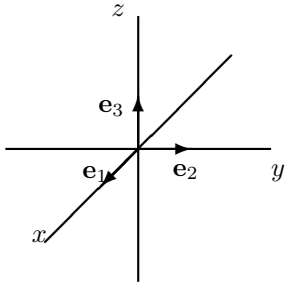
Thus the magnitude of $r\mathbf{v}$ equals $|r|$ times the magnitude of $\mathbf{v}$. If r is positive, then the vector represented by $r\mathbf{v}$ has the same direction as the vector, $\mathbf{v}$ because multiplying by the scalar, r, only has the effect of scaling all the distances. Thus the unit distance along any coordinate axis now has length r and in this rescaled system the vector is represented by $\mathbf{a}$. If $r < 0$ similar considerations apply except in this case all the a_i also change sign. From now on, $\mathbf{a}$ will be referred to as a vector instead of an element of $\mathbb{R}^n$ representing a vector as just described. The following picture illustrates the effect of scalar multiplication.



Note there are n special vectors which point along the coordinate axes. These are

$$\mathbf{e}_i \equiv (0,\cdots,0,1,0,\cdots,0)$$

where the 1 is in the i^{th} slot and there are zeros in all the other spaces. See the picture in the case of $\mathbb{R}^3$.



The direction of $\mathbf{e}_i$ is referred to as the i^{th} direction. Given a vector, $\mathbf{v} = (a_1, \cdots, a_n)$, $a_i\mathbf{e}_i$ is the i^{th} component of the vector. Thus $a_i\mathbf{e}_i = (0, \cdots, 0, a_i, 0, \cdots, 0)$ and so this vector gives something possibly nonzero only in the i^{th} direction. Also, knowledge of the i^{th} component of the vector is equivalent to knowledge of the vector because it gives the entry in the i^{th} slot and for $\mathbf{v} = (a_1, \cdots, a_n)$,

$$\mathbf{v} = \sum_{k=1}^{n} a_i\mathbf{e}_i.$$

What does addition of vectors mean physically? Suppose two forces are applied to some object. Each of these would be represented by a force vector and the two forces acting together would yield an overall force acting on the object which would also be a force vector known as the resultant. Suppose the two vectors are $\mathbf{a} = \sum_{k=1}^{n} a_i\mathbf{e}_i$ and $\mathbf{b} = \sum_{k=1}^{n} b_i\mathbf{e}_i$. Then the vector, $\mathbf{a}$ involves a component in the i^{th} direction, $a_i\mathbf{e}_i$ while the component in the i^{th} direction of $\mathbf{b}$ is $b_i\mathbf{e}_i$. Then it seems physically reasonable that the resultant vector should have a component in the i^{th} direction equal to $(a_i + b_i)\,\mathbf{e}_i$. This is exactly what is obtained when the vectors, $\mathbf{a}$ and $\mathbf{b}$ are added.

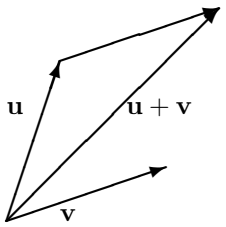
$$\begin{aligned}\mathbf{a} + \mathbf{b} &= (a_1 + b_1, \cdots, a_n + b_n)\,. \\ &= \sum_{i=1}^{n} (a_i + b_i)\,\mathbf{e}_i.\end{aligned}$$

Thus the addition of vectors according to the rules of addition in $\mathbb{R}^n$ which were presented earlier, yields the appropriate vector which duplicates the cumulative effect of all the vectors in the sum.

What is the geometric significance of vector addition? Suppose $\mathbf{u}, \mathbf{v}$ are vectors,

$$\mathbf{u} = (u_1, \cdots, u_n)\,, \mathbf{v} = (v_1, \cdots, v_n)$$

Then $\mathbf{u} + \mathbf{v} = (u_1 + v_1, \cdots, u_n + v_n)$. How can one obtain this geometrically? Consider the directed line segment, $\overrightarrow{\mathbf{0u}}$ and then, starting at the end of this directed line segment, follow the directed line segment $\overrightarrow{\mathbf{u}(\mathbf{u}+\mathbf{v})}$ to its end, $\mathbf{u} + \mathbf{v}$. In other words, place the vector $\mathbf{u}$ in standard position with its base at the origin and then slide the vector $\mathbf{v}$ till its base coincides with the point of $\mathbf{u}$. The point of this slid vector, determines $\mathbf{u} + \mathbf{v}$. To illustrate, see the following picture



Note the vector $\mathbf{u} + \mathbf{v}$ is the diagonal of a parallelogram determined from the two vectors $\mathbf{u}$ and $\mathbf{v}$ and that identifying $\mathbf{u} + \mathbf{v}$ with the directed diagonal of the parallelogram determined by the vectors $\mathbf{u}$ and $\mathbf{v}$ amounts to the same thing as the above procedure.

An item of notation should be mentioned here. In the case of $\mathbb{R}^n$ where $n \leq 3$, it is standard notation to use $\mathbf{i}$ for $\mathbf{e}_1, \mathbf{j}$ for $\mathbf{e}_2$, and $\mathbf{k}$ for $\mathbf{e}_3$. Now here are some applications of vector addition to some problems.

Example 2.6.4 *There are three ropes attached to a car and three people pull on these ropes. The first exerts a force of* $2\mathbf{i}+3\mathbf{j}-2\mathbf{k}$ *Newtons, the second exerts a force of* $3\mathbf{i}+5\mathbf{j}+\mathbf{k}$ *Newtons*

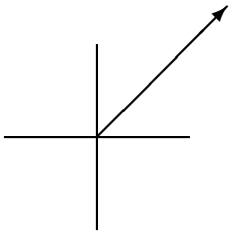
and the third exerts a force of $5\mathbf{i}-\mathbf{j}+2\mathbf{k}$. *Newtons. Find the total force in the direction of* $\mathbf{i}$.

To find the total force add the vectors as described above. This gives $10\mathbf{i}+7\mathbf{j}+\mathbf{k}$ Newtons. Therefore, the force in the $\mathbf{i}$ direction is 10 Newtons.

As mentioned earlier, the Newton is a unit of force like pounds.

Example 2.6.5 *An airplane flies North East at 100 miles per hour. Write this as a vector.*

A picture of this situation follows.



The vector has length 100. Now using that vector as the hypotenuse of a right triangle having equal sides, the sides should be each of length $100/\sqrt{2}$. Therefore, the vector would be $100/\sqrt{2}\mathbf{i}+100/\sqrt{2}\mathbf{j}$.

This example also motivates the concept of **velocity**.

Definition 2.6.6 *The* ***speed*** *of an object is a measure of how fast it is going. It is measured in units of length per unit time. For example, miles per hour, kilometers per minute, feet per second. The* ***velocity*** *is a vector having the speed as the magnitude but also specifying the direction.*

Thus the velocity vector in the above example is $100/\sqrt{2}\mathbf{i}+100/\sqrt{2}\mathbf{j}$.

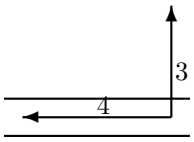
Example 2.6.7 *The velocity of an airplane is* $100\mathbf{i}+\mathbf{j}+\mathbf{k}$ *measured in kilometers per hour and at a certain instant of time its position is* $(1,2,1)$. *Here imagine a Cartesian coordinate system in which the third component is altitude and the first and second components are measured on a line from West to East and a line from South to North. Find the position of this airplane one minute later.*

Consider the vector $(1,2,1)$, is the initial position vector of the airplane. As it moves, the position vector changes. After one minute the airplane has moved in the $\mathbf{i}$ direction a distance of $100\times\frac{1}{60}=\frac{5}{3}$ kilometer. In the $\mathbf{j}$ direction it has moved $\frac{1}{60}$ kilometer during this same time, while it moves $\frac{1}{60}$ kilometer in the $\mathbf{k}$ direction. Therefore, the new displacement vector for the airplane is

$$(1,2,1)+\left(\frac{5}{3},\frac{1}{60},\frac{1}{60}\right)=\left(\frac{8}{3},\frac{121}{60},\frac{121}{60}\right)$$

Example 2.6.8 *A certain river is one half mile wide with a current flowing at 4 miles per hour from East to West. A man swims directly toward the opposite shore from the South bank of the river at a speed of 3 miles per hour. How far down the river does he find himself when he has swam across? How far does he end up swimming?*

Consider the following picture.



You should write these vectors in terms of components. The velocity of the swimmer in still water would be $3\mathbf{j}$ while the velocity of the river would be $-4\mathbf{i}$. Therefore, the velocity of the swimmer is $-4\mathbf{i}+3\mathbf{j}$. Since the component of velocity in the direction across the river is 3, it follows the trip takes 1/6 hour or 10 minutes. The speed at which he travels is $\sqrt{4^2+3^2}=5$ miles per hour and so he travels $5\times\frac{1}{6}=\frac{5}{6}$ miles. Now to find the distance downstream he finds himself, note that if x is this distance, x and $1/2$ are two legs of a right triangle whose hypotenuse equals 5/6 miles. Therefore, by the Pythagorean theorem the distance downstream is

$$\sqrt{(5/6)^2-(1/2)^2}=\frac{2}{3}\text{ miles.}$$

2.7 Exercises With Answers

1. The wind blows from West to East at a speed of 30 kilometers per hour and an airplane which travels at 300 Kilometers per hour in still air is heading North West. What is the velocity of the airplane relative to the ground? What is the component of this velocity in the direction North?

   Let the positive y axis point in the direction North and let the positive x axis point in the direction East. The velocity of the wind is $30\mathbf{i}$. The plane moves in the direction $\mathbf{i}+\mathbf{j}$. A unit vector in this direction is $\frac{1}{\sqrt{2}}(\mathbf{i}+\mathbf{j})$. Therefore, the velocity of the plane relative to the ground is $30\mathbf{i}+\frac{300}{\sqrt{2}}(\mathbf{i}+\mathbf{j})=150\sqrt{2}\mathbf{j}+\left(30+150\sqrt{2}\right)\mathbf{i}$. The component of velocity in the direction North is $150\sqrt{2}$.

2. In the situation of Problem 1 how many degrees to the West of North should the airplane head in order to fly exactly North. What will be the speed of the airplane relative to the ground?

   In this case the unit vector will be $-\sin(\theta)\mathbf{i}+\cos(\theta)\mathbf{j}$. Therefore, the velocity of the plane will be

   $$300\left(-\sin(\theta)\mathbf{i}+\cos(\theta)\mathbf{j}\right)$$

   and this is supposed to satisfy

   $$300\left(-\sin(\theta)\mathbf{i}+\cos(\theta)\mathbf{j}\right)+30\mathbf{i}=0\mathbf{i}+?\mathbf{j}.$$

   Therefore, you need to have $\sin\theta=1/10$, which means $\theta=.100\,17$ radians. Therefore, the degrees should be $\frac{.1\times180}{\pi}=5.729\,6$ degrees. In this case the velocity vector of the plane relative to the ground is $300\left(\frac{\sqrt{99}}{10}\right)\mathbf{j}$.

3. In the situation of 2 suppose the airplane uses 34 gallons of fuel every hour at that air speed and that it needs to fly North a distance of 600 miles. Will the airplane have enough fuel to arrive at its destination given that it has 63 gallons of fuel?

   The airplane needs to fly 600 miles at a speed of $300\left(\frac{\sqrt{99}}{10}\right)$. Therefore, it takes $\frac{600}{\left(300\left(\frac{\sqrt{99}}{10}\right)\right)}=2.010\,1$ hours to get there. Therefore, the plane will need to use about 68 gallons of gas. It won't make it.

4. A certain river is one half mile wide with a current flowing at 3 miles per hour from East to West. A man swims directly toward the opposite shore from the South bank of the river at a speed of 2 miles per hour. How far down the river does he find himself when he has swam across? How far does he end up swimming?

   The velocity of the man relative to the earth is then $-3\mathbf{i}+2\mathbf{j}$. Since the component of $\mathbf{j}$ equals 2 it follows he takes 1/8 of an hour to get across. During this time he is swept downstream at the rate of 3 miles per hour and so he ends up 3/8 of a mile down stream. He has gone $\sqrt{\left(\frac{3}{8}\right)^2+\left(\frac{1}{2}\right)^2}=.625$ miles in all.

5. Three forces are applied to a point which does not move. Two of the forces are $2\mathbf{i}-\mathbf{j}+3\mathbf{k}$ Newtons and $\mathbf{i}-3\mathbf{j}-2\mathbf{k}$ Newtons**.** Find the third force.

   Call it $a\mathbf{i}+b\mathbf{j}+c\mathbf{k}$ Then you need $a+2+1=0, b-1-3=0$, and $c+3-2=0$. Therefore, the force is $-3\mathbf{i}+4\mathbf{j}-\mathbf{k}$.

Systems Of Equations

3.0.1 Outcomes

A. Relate the types of solution sets of a system of two or three variables to the intersections of lines in a plane or the intersection of planes in three space.

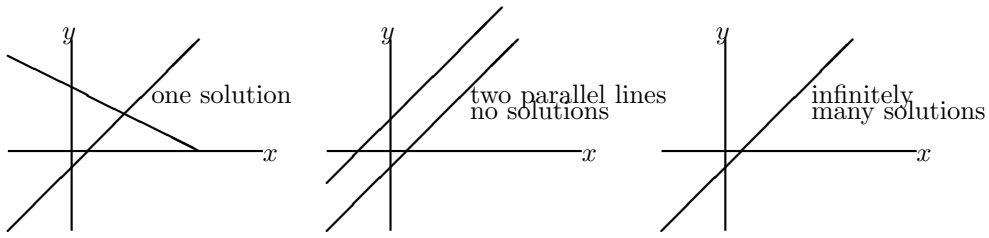
B. Determine whether a system of linear equations has no solution, a unique solution or an infinite number of solutions from its echelon form.

C. Solve a system of equations using Gauss elimination.

D. Model a physical system with linear equations and then solve.

3.1 Systems Of Equations, Geometric Interpretations

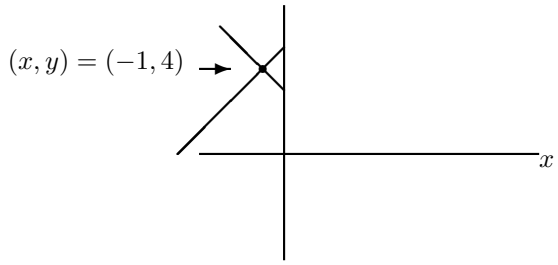
As you know, equations like $2x + 3y = 6$ can be graphed as straight lines in $\mathbb{R}^2$. To find the solution to two such equations, you could graph the two straight lines and the ordered pairs identifying the point (or points) of intersection would give the x and y values of the solution to the two equations because such an ordered pair satisfies both equations. The following picture illustrates what can occur with two equations involving two variables.



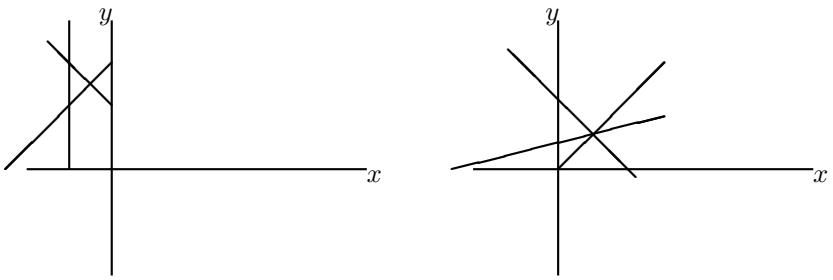
In the first example of the above picture, there is a unique point of intersection. In the second, there are no points of intersection. The other thing which can occur is that the two lines are really the same line. For example, $x + y = 1$ and $2x + 2y = 2$ are relations which when graphed yield the same line. In this case there are infinitely many points in the simultaneous solution of these two equations, every ordered pair which is on the graph of the line. It is always this way when considering linear systems of equations. There is either no solution, exactly one or infinitely many although the reasons for this are not completely comprehended by considering a simple picture in two dimensions, $\mathbb{R}^2$.

Example 3.1.1 *Find the solution to the system* $x + y = 3$, $y - x = 5$.

You can verify the solution is $(x, y) = (-1, 4)$. You can see this geometrically by graphing the equations of the two lines. If you do so correctly, you should obtain a graph which looks something like the following in which the point of intersection represents the solution of the two equations.

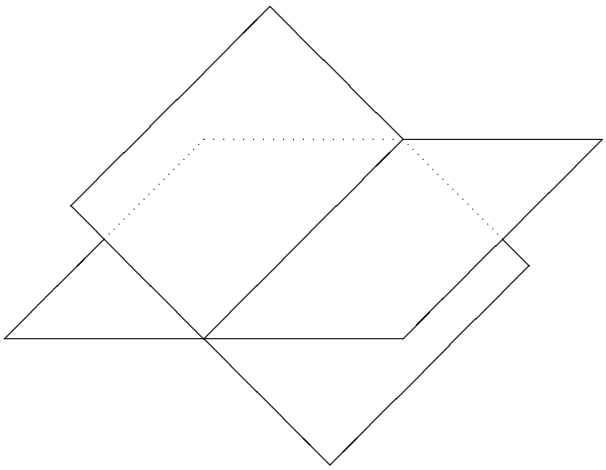


Example 3.1.2 *You can also imagine other situations such as the case of three intersecting lines having no common point of intersection or three intersecting lines which do intersect at a single point as illustrated in the following picture.*



In the case of the first picture above, there would be no solution to the three equations whose graphs are the given lines. In the case of the second picture there is a solution to the three equations whose graphs are the given lines.

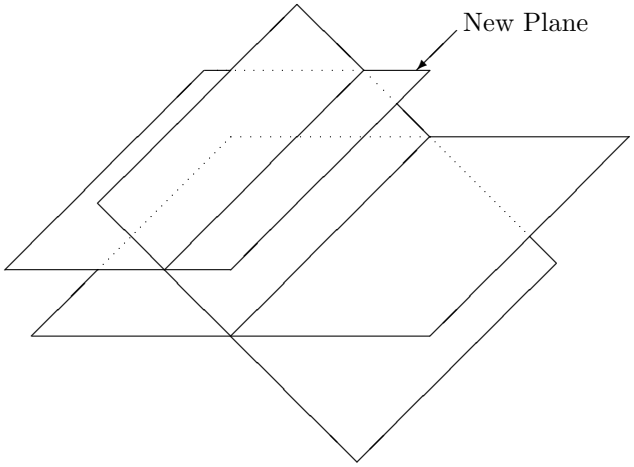
The points, (x, y, z) satisfying an equation in three variables like $2x + 4y - 5z = 8$ form a plane [1] and geometrically, when you solve systems of equations involving three variables, you are taking intersections of planes. Consider the following picture involving two planes.



[1] Don't worry about why this is at this time. It is not important. The following discussion is intended to show you that geometric considerations like this don't take you anywhere. It is the algebraic procedures which are important and lead to important applications.

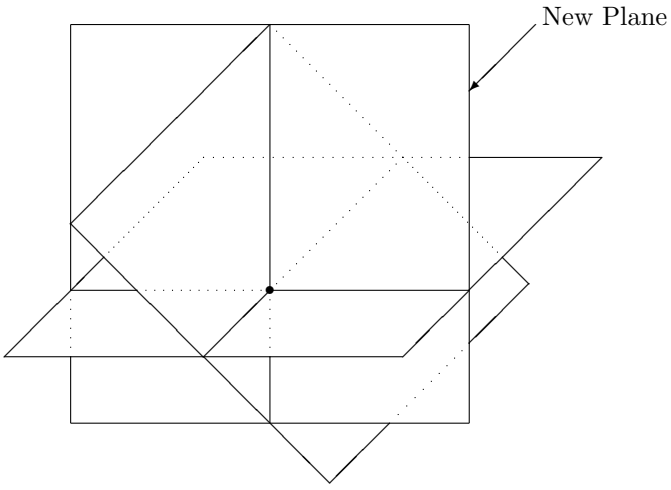
Notice how these two planes intersect in a line. It could also happen the two planes could fail to intersect.

Now imagine a third plane. One thing that could happen is this third plane could have an intersection with one of the first planes which results in a line which fails to intersect the first line as illustrated in the following picture.



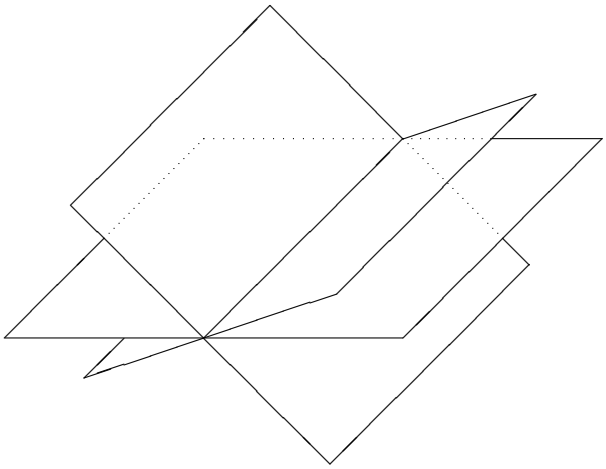
Thus there is no point which lies in all three planes. The picture illustrates the situation in which the line of intersection of the new plane with one of the original planes forms a line parallel to the line of intersection of the first two planes. However, in three dimensions, it is possible for two lines to fail to intersect even though they are not parallel. Such lines are called **skew lines.** You might consider whether there exist two skew lines, each of which is the intersection of a pair of planes selected from a set of exactly three planes such that there is no point of intersection between the three planes. You can also see that if you tilt one of the planes you could obtain every pair of planes having a nonempty intersection in a line and yet there may be no point in the intersection of all three.

It could happen also that the three planes could intersect in a single point as shown in the following picture.



In this case, the three planes have a single point of intersection. The three planes could

also intersect in a line.



Thus in the case of three equations having three variables, the planes determined by these equations could intersect in a single point, a line, or even fail to intersect at all. You see that in three dimensions there are many possibilities. If you want to waste some time, you can try to imagine all the things which could happen but this will not help for more variables than 3 which is where many of the important applications lie.

Relations like $x + y - 2z + 4w = 8$ are often called **hyper-planes**.[2] However, it is impossible to draw pictures of such things.The only rational and useful way to deal with this subject is through the use of algebra not art. Mathematics exists partly to free us from having to always draw pictures in order to draw conclusions.

3.2 Systems Of Equations, Algebraic Procedures

3.2.1 Elementary Operations

Consider the following example.

Example 3.2.1 *Find x and y such that*

$$x + y = 7 \text{ and } 2x - y = 8. \tag{3.1}$$

*The set of ordered pairs, (x, y) which solve both equations is called the **solution set**.*

You can verify that $(x, y) = (5, 2)$ is a solution to the above system. The interesting question is this: If you were not given this information to verify, how could you determine the solution? You can do this by using the following basic operations on the equations, none of which change the set of solutions of the system of equations.

Definition 3.2.2 ***Elementary operations*** *are those operations consisting of the following.*

1. *Interchange the order in which the equations are listed.*

[2]The evocative semi word, "hyper" conveys absolutely no meaning but is traditional usage which makes the terminology sound more impressive than something like long wide flat thing.Later we will discuss some terms which are not just evocative but yield real understanding.

2. *Multiply any equation by a nonzero number.*

3. *Replace any equation with itself added to a multiple of another equation.*

Example 3.2.3 *To illustrate the third of these operations on this particular system, consider the following.*

$$\begin{array}{c} x+y=7 \\ 2x-y=8 \end{array}$$

The system has the same solution set as the system

$$\begin{array}{c} x+y=7 \\ -3y=-6 \end{array}.$$

To obtain the second system, take the second equation of the first system and add -2 times the first equation to obtain

$$-3y=-6.$$

Now, this clearly shows that $y=2$ and so it follows from the other equation that $x+2=7$ and so $x=5$.

Of course a linear system may involve many equations and many variables. The solution set is still the collection of solutions to the equations. In every case, the above operations of Definition 3.2.2 do not change the set of solutions to the system of linear equations.

Theorem 3.2.4 *Suppose you have two equations, involving the variables,* $(x_1,\cdots,x_n)$

$$E_1=f_1, E_2=f_2 \tag{3.2}$$

where E_1 *and* E_2 *are expressions involving the variables and* f_1 *and* f_2 *are constants. (In the above example there are only two variables,* x *and* y *and* $E_1=x+y$ *while* $E_2=2x-y$.*) Then the system* $E_1=f_1, E_2=f_2$ *has the same solution set as*

$$E_1=f_1,\ E_2+aE_1=f_2+af_1. \tag{3.3}$$

Also the system $E_1=f_1, E_2=f_2$ *has the same solutions as the system,* $E_2=f_2, E_1=f_1$. *The system* $E_1=f_1, E_2=f_2$ *has the same solution as the system* $E_1=f_1, aE_2=af_2$ *provided* $a\neq 0$.

Proof: If $(x_1,\cdots,x_n)$ solves $E_1=f_1, E_2=f_2$ then it solves the first equation in $E_1=f_1,\ E_2+aE_1=f_2+af_1$. Also, it satisfies $aE_1=af_1$ and so, since it also solves $E_2=f_2$ it must solve $E_2+aE_1=f_2+af_1$. Therefore, if $(x_1,\cdots,x_n)$ solves $E_1=f_1, E_2=f_2$ it must also solve $E_2+aE_1=f_2+af_1$. On the other hand, if it solves the system $E_1=f_1$ and $E_2+aE_1=f_2+af_1$, then $aE_1=af_1$ and so you can subtract these equal quantities from both sides of $E_2+aE_1=f_2+af_1$ to obtain $E_2=f_2$ showing that it satisfies $E_1=f_1, E_2=f_2$.

The second assertion of the theorem which says that the system $E_1=f_1, E_2=f_2$ has the same solution as the system, $E_2=f_2, E_1=f_1$ is seen to be true because it involves nothing more than listing the two equations in a different order. They are the same equations.

The third assertion of the theorem which says $E_1=f_1, E_2=f_2$ has the same solution as the system $E_1=f_1, aE_2=af_2$ provided $a\neq 0$ is verified as follows: If $(x_1,\cdots,x_n)$ is a solution of $E_1=f_1, E_2=f_2$, then it is a solution to $E_1=f_1, aE_2=af_2$ because the second system only involves multiplying the equation, $E_2=f_2$ by a. If $(x_1,\cdots,x_n)$ is a solution of $E_1=f_1, aE_2=af_2$, then upon multiplying $aE_2=af_2$ by the number, $1/a$, you find that $E_2=f_2$.

Stated simply, the above theorem shows that the elementary operations do not change the solution set of a system of equations.

Here is an example in which there are three equations and three variables. You want to find values for x, y, z such that each of the given equations are satisfied when these values are plugged in to the equations.

Example 3.2.5 *Find the solutions to the system,*

$$\begin{array}{c} x+3y+6z=25 \\ 2x+7y+14z=58 \\ 2y+5z=19 \end{array} \tag{3.4}$$

To solve this system replace the second equation by (-2) times the first equation added to the second. This yields the system

$$\begin{array}{c} x+3y+6z=25 \\ y+2z=8 \\ 2y+5z=19 \end{array} \tag{3.5}$$

Now take (-2) times the second and add to the third. More precisely, replace the third equation with (-2) times the second added to the third. This yields the system

$$\begin{array}{c} x+3y+6z=25 \\ y+2z=8 \\ z=3 \end{array} \tag{3.6}$$

At this point, you can tell what the solution is. This system has the same solution as the original system and in the above, $z=3$. Then using this in the second equation, it follows $y+6=8$ and so $y=2$. Now using this in the top equation yields $x+6+18=25$ and so $x=1$. This process is called **back substitution**.

Alternatively, in 3.6 you could have continued as follows. Add (-2) times the bottom equation to the middle and then add (-6) times the bottom to the top. This yields

$$\begin{array}{c} x+3y=7 \\ y=2 \\ z=3 \end{array}$$

Now add (-3) times the second to the top. This yields

$$\begin{array}{c} x=1 \\ y=2 \\ z=3 \end{array},$$

a system which has the same solution set as the original system. This avoided back substitution and led to the same solution set.

3.2.2 Gauss Elimination

A less cumbersome way to represent a linear system is to write it as an **augmented matrix.** For example the linear system, 3.4 can be written as

$$\left(\begin{array}{ccc|c} 1 & 3 & 6 & 25 \\ 2 & 7 & 14 & 58 \\ 0 & 2 & 5 & 19 \end{array}\right).$$

It has exactly the same information as the original system but here it is understood there is an x column, $\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$, a y column, $\begin{pmatrix} 3 \\ 7 \\ 2 \end{pmatrix}$ and a z column, $\begin{pmatrix} 6 \\ 14 \\ 5 \end{pmatrix}$. The rows correspond

to the equations in the system. Thus the top row in the augmented matrix corresponds to the equation,

$$x + 3y + 6z = 25.$$

Now when you replace an equation with a multiple of another equation added to itself, you are just taking a row of this augmented matrix and replacing it with a multiple of another row added to it. Thus the first step in solving 3.4 would be to take (-2) times the first row of the augmented matrix above and add it to the second row,

$$\left(\begin{array}{ccc|c} 1 & 3 & 6 & 25 \\ 0 & 1 & 2 & 8 \\ 0 & 2 & 5 & 19 \end{array}\right).$$

Note how this corresponds to 3.5. Next take (-2) times the second row and add to the third,

$$\left(\begin{array}{ccc|c} 1 & 3 & 6 & 25 \\ 0 & 1 & 2 & 8 \\ 0 & 0 & 1 & 3 \end{array}\right)$$

This augmented matrix corresponds to the system

$$\begin{array}{c} x + 3y + 6z = 25 \\ y + 2z = 8 \\ z = 3 \end{array}$$

which is the same as 3.6. By back substitution you obtain the solution $x = 1, y = 6$, and $z = 3$.

In general a linear system is of the form

$$\begin{array}{c} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m \end{array}, \tag{3.7}$$

where the x_i are variables and the a_{ij} and b_i are constants. This system can be represented by the augmented matrix,

$$\left(\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{array}\right). \tag{3.8}$$

Changes to the system of equations in 3.7 as a result of an elementary operations translate into changes of the augmented matrix resulting from a row operation. Note that Theorem 3.2.4 implies that the row operations deliver an augmented matrix for a system of equations which has the same solution set as the original system.

Definition 3.2.6 *The **row operations** consist of the following*

1. *Switch two rows.*
2. *Multiply a row by a nonzero number.*
3. *Replace a row by a multiple of another row added to it.*

Gauss elimination is a systematic procedure to simplify an augmented matrix to a reduced form. In the following definition, the term "**leading entry**" refers to the first nonzero entry of a row when scanning the row from left to right.

Definition 3.2.7 *An augmented matrix is in* ***echelon form*** *if*

1. *All nonzero rows are above any rows of zeros.*
2. *Each leading entry of a row is in a column to the right of the leading entries of any rows above it.*

Definition 3.2.8 *An augmented matrix is in* ***row reduced echelon form*** *if*

1. *All nonzero rows are above any rows of zeros.*
2. *Each leading entry of a row is in a column to the right of the leading entries of any rows above it.*
3. *All entries in a column above and below a leading entry are zero.*
4. *Each leading entry is a 1, the only nonzero entry in its column.*

Example 3.2.9 *Here are some augmented matrices which are in row reduced echelon form.*

$$\left(\begin{array}{ccccc|c} 1 & 0 & 0 & 5 & 8 & 0 \\ 0 & 0 & 1 & 2 & 7 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right), \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

Example 3.2.10 *Here are augmented matrices in echelon form which are not in row reduced echelon form but which are in echelon form.*

$$\left(\begin{array}{ccccc|c} 1 & 0 & 6 & 5 & 8 & 2 \\ 0 & 0 & 2 & 2 & 7 & 3 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right), \left(\begin{array}{ccc|c} 1 & 3 & 5 & 4 \\ 0 & 2 & 0 & 7 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Example 3.2.11 *Here are some augmented matrices which are not in echelon form.*

$$\left(\begin{array}{ccc|c} 0 & 0 & 0 & 0 \\ 1 & 2 & 3 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right), \left(\begin{array}{cc|c} 1 & 2 & 3 \\ 2 & 4 & -6 \\ 4 & 0 & 7 \end{array}\right), \left(\begin{array}{ccc|c} 0 & 2 & 3 & 3 \\ 1 & 5 & 0 & 2 \\ 7 & 5 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{array}\right).$$

Definition 3.2.12 *A* ***pivot position*** *in a matrix is the location of a leading entry in an echelon form resulting from the application of row operations to the matrix. A* ***pivot column*** *is a column that contains a pivot position.*

For example consider the following.

Example 3.2.13 *Suppose*

$$A = \left(\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 6 \\ 4 & 4 & 4 & 10 \end{array}\right)$$

Where are the pivot positions and pivot columns?

Replace the second row by -3 times the first added to the second. This yields

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -6 \\ 4 & 4 & 4 & 10 \end{array}\right).$$

This is not in reduced echelon form so replace the bottom row by -4 times the top row added to the bottom. This yields

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -6 \\ 0 & -4 & -8 & -6 \end{array}\right).$$

This is still not in reduced echelon form. Replace the bottom row by -1 times the middle row added to the bottom. This yields

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 4 \\ 0 & -4 & -8 & -6 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

which is in echelon form, although not in reduced echelon form. Therefore, the pivot positions in the original matrix are the locations corresponding to the first row and first column and the second row and second columns as shown in the following:

$$\left(\begin{array}{ccc|c} \boxed{1} & 2 & 3 & 4 \\ 3 & \boxed{2} & 1 & 6 \\ 4 & 4 & 4 & 10 \end{array}\right)$$

Thus the pivot columns in the matrix are the first two columns.

The following is the algorithm for obtaining a matrix which is in row reduced echelon form.

Algorithm 3.2.14

This algorithm tells how to start with a matrix and do row operations on it in such a way as to end up with a matrix in row reduced echelon form.

1. Find the first nonzero column from the left. This is the first pivot column. The position at the top of the first pivot column is the first pivot position. Switch rows if necessary to place a nonzero number in the first pivot position.

2. Use row operations to zero out the entries below the first pivot position.

3. Ignore the row containing the most recent pivot position identified and the rows above it. Repeat steps 1 and 2 to the remaining sub-matrix, the rectangular array of numbers obtained from the original matrix by deleting the rows you just ignored. Repeat the process until there are no more rows to modify. The matrix will then be in echelon form.

4. Moving from right to left, use the nonzero elements in the pivot positions to zero out the elements in the pivot columns which are above the pivots.

5. Divide each nonzero row by the value of the leading entry. The result will be a matrix in row reduced echelon form.

This row reduction procedure applies to both augmented matrices and non augmented matrices. There is nothing special about the augmented column with respect to the row reduction procedure.

Example 3.2.15 *Here is a matrix.*

$$\begin{pmatrix} 0 & 0 & 2 & 3 & 2 \\ 0 & 1 & 1 & 4 & 3 \\ 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 \end{pmatrix}$$

Do row reductions till you obtain a matrix in echelon form. Then complete the process by producing one in reduced echelon form.

The pivot column is the second. Hence the pivot position is the one in the first row and second column. Switch the first two rows to obtain a nonzero entry in this pivot position.

$$\begin{pmatrix} 0 & 1 & 1 & 4 & 3 \\ 0 & 0 & 2 & 3 & 2 \\ 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 \end{pmatrix}$$

Step two is not necessary because all the entries below the first pivot position in the resulting matrix are zero. Now ignore the top row and the columns to the left of this first pivot position. Thus you apply the same operations to the smaller matrix,

$$\begin{pmatrix} 2 & 3 & 2 \\ 1 & 2 & 2 \\ 0 & 0 & 0 \\ 0 & 2 & 1 \end{pmatrix}.$$

The next pivot column is the third corresponding to the first in this smaller matrix and the second pivot position is therefore, the one which is in the second row and third column. In this case it is not necessary to switch any rows to place a nonzero entry in this position because there is already a nonzero entry there. Multiply the third row of the original matrix by -2 and then add the second row to it. This yields

$$\begin{pmatrix} 0 & 1 & 1 & 4 & 3 \\ 0 & 0 & 2 & 3 & 2 \\ 0 & 0 & 0 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 \end{pmatrix}.$$

The next matrix the steps in the algorithm are applied to is

$$\begin{pmatrix} -1 & -2 \\ 0 & 0 \\ 2 & 1 \end{pmatrix}.$$

The first pivot column is the first column in this case and no switching of rows is necessary because there is a nonzero entry in the first pivot position. Therefore, the algorithm yields

for the next step

$$\begin{pmatrix} 0 & 1 & 1 & 4 & 3 \\ 0 & 0 & 2 & 3 & 2 \\ 0 & 0 & 0 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -3 \end{pmatrix}.$$

Now the algorithm will be applied to the matrix,

$$\begin{pmatrix} 0 \\ -3 \end{pmatrix}$$

There is only one column and it is nonzero so this single column is the pivot column. Therefore, the algorithm yields the following matrix for the echelon form.

$$\begin{pmatrix} 0 & 1 & 1 & 4 & 3 \\ 0 & 0 & 2 & 3 & 2 \\ 0 & 0 & 0 & -1 & -2 \\ 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

To complete placing the matrix in reduced echelon form, multiply the third row by 3 and add -2 times the fourth row to it. This yields

$$\begin{pmatrix} 0 & 1 & 1 & 4 & 3 \\ 0 & 0 & 2 & 3 & 2 \\ 0 & 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Next multiply the second row by 3 and take 2 times the fourth row and add to it. Then add the fourth row to the first.

$$\begin{pmatrix} 0 & 1 & 1 & 4 & 0 \\ 0 & 0 & 6 & 9 & 0 \\ 0 & 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Next work on the fourth column in the same way.

$$\begin{pmatrix} 0 & 3 & 3 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 \\ 0 & 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Take $-1/2$ times the second row and add to the first.

$$\begin{pmatrix} 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 6 & 0 & 0 \\ 0 & 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & 0 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Finally, divide by the value of the leading entries in the nonzero rows.

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

The above algorithm is the way a computer would obtain a reduced echelon form for a given matrix. It is not necessary for you to pretend you are a computer but if you like to do so, the algorithm described above will work. The main idea is to do row operations in such a way as to end up with a matrix in echelon form or row reduced echelon form because when this has been done, the resulting augmented matrix will allow you to describe the solutions to the linear system of equations in a meaningful way.

Example 3.2.16 *Give the complete solution to the system of equations,* $5x+10y-7z=-2$, $2x+4y-3z=-1$, *and* $3x+6y+5z=9$.

The augmented matrix for this system is

$$\left(\begin{array}{ccc|c} 2 & 4 & -3 & -1 \\ 5 & 10 & -7 & -2 \\ 3 & 6 & 5 & 9 \end{array}\right)$$

Multiply the second row by 2, the first row by 5, and then take (-1) times the first row and add to the second. Then multiply the first row by $1/5$. This yields

$$\left(\begin{array}{ccc|c} 2 & 4 & -3 & -1 \\ 0 & 0 & 1 & 1 \\ 3 & 6 & 5 & 9 \end{array}\right)$$

Now, combining some row operations, take (-3) times the first row and add this to 2 times the last row and replace the last row with this. This yields.

$$\left(\begin{array}{ccc|c} 2 & 4 & -3 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 21 \end{array}\right).$$

One more row operation, taking (-1) times the second row and adding to the bottom yields.

$$\left(\begin{array}{ccc|c} 2 & 4 & -3 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 20 \end{array}\right).$$

This is impossible because the last row indicates the need for a solution to the equation

$$0x+0y+0z=20$$

and there is no such thing because $0 \neq 20$. This shows there is no solution to the three given equations. When this happens, the system is called **inconsistent**. In this case it is very easy to describe the solution set. The system has no solution.

Here is another example based on the use of row operations.

Example 3.2.17 *Give the complete solution to the system of equations,* $3x-y-5z=9$, $y-10z=0$, *and* $-2x+y=-6$.

The augmented matrix of this system is

$$\left(\begin{array}{ccc|c} 3 & -1 & -5 & 9 \\ 0 & 1 & -10 & 0 \\ -2 & 1 & 0 & -6 \end{array}\right)$$

Replace the last row with 2 times the top row added to 3 times the bottom row. This gives

$$\left(\begin{array}{ccc|c} 3 & -1 & -5 & 9 \\ 0 & 1 & -10 & 0 \\ 0 & 1 & -10 & 0 \end{array}\right).$$

The entry, 3 in this sequence of row operations is called the **pivot**. It is used to create zeros in the other places of the column. Next take -1 times the middle row and add to the bottom. Here the 1 in the second row is the pivot.

$$\left(\begin{array}{ccc|c} 3 & -1 & -5 & 9 \\ 0 & 1 & -10 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Take the middle row and add to the top and then divide the top row which results by 3.

$$\left(\begin{array}{ccc|c} 1 & 0 & -5 & 3 \\ 0 & 1 & -10 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

This is in reduced echelon form. The equations corresponding to this reduced echelon form are $y = 10z$ and $x = 3 + 5z$. Apparently z can equal any number. Lets call this number, t. [3]Therefore, the solution set of this system is $x = 3 + 5t, y = 10t$, and $z = t$ where t is completely arbitrary. The system has an infinite set of solutions which are given in the above simple way. This is what it is all about, finding the solutions to the system.

There is some terminology connected to this which is useful. Recall how each column corresponds to a variable in the original system of equations. The variables corresponding to a pivot column are called **basic variables** . The other variables are called **free variables.** In Example 3.2.17 there was one free variable, z, and two basic variables, x and y. In describing the solution to the system of equations, the free variables are assigned a parameter. In Example 3.2.17 this parameter was t. Sometimes there are many free variables and in these cases, you need to use many parameters. Here is another example.

Example 3.2.18 *Find the solution to the system*

$$\begin{aligned} x + 2y - z + w &= 3 \\ x + y - z + w &= 1 \\ x + 3y - z + w &= 5 \end{aligned}$$

The augmented matrix is

$$\left(\begin{array}{cccc|c} 1 & 2 & -1 & 1 & 3 \\ 1 & 1 & -1 & 1 & 1 \\ 1 & 3 & -1 & 1 & 5 \end{array}\right).$$

Take -1 times the first row and add to the second. Then take -1 times the first row and add to the third. This yields

$$\left(\begin{array}{cccc|c} 1 & 2 & -1 & 1 & 3 \\ 0 & -1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 0 & 2 \end{array}\right)$$

[3]In this context t is called a **parameter.**

Now add the second row to the bottom row

$$\left(\begin{array}{cccc|c} 1 & 2 & -1 & 1 & 3 \\ 0 & -1 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right) \tag{3.9}$$

This matrix is in echelon form and you see the basic variables are x and y while the free variables are z and w. Assign s to z and t to w. Then the second row yields the equation, $y = 2$ while the top equation yields the equation, $x + 2y - s + t = 3$ and so since $y = 2$, this gives $x + 4 - s + t = 3$ showing that $x = -1 + s - t, y = 2, z = s$, and $w = t$. It is customary to write this in the form

$$\left(\begin{array}{c} x \\ y \\ z \\ w \end{array}\right) = \left(\begin{array}{c} -1 + s - t \\ 2 \\ s \\ t \end{array}\right). \tag{3.10}$$

This is another example of a system which has an infinite solution set but this time the solution set depends on two parameters, not one. Most people find it less confusing in the case of an infinite solution set to first place the augmented matrix in row reduced echelon form rather than just echelon form before seeking to write down the description of the solution. In the above, this means we don't stop with the echelon form 3.9. Instead we first place it in reduced echelon form as follows.

$$\left(\begin{array}{cccc|c} 1 & 0 & -1 & 1 & -1 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Then the solution is $y = 2$ from the second row and $x = -1 + z - w$ from the first. Thus letting $z = s$ and $w = t$, the solution is given in 3.10.

The number of free variables is always equal to the number of **different** parameters used to describe the solution. If there are no free variables, then either there is no solution as in the case where row operations yield an echelon form like

$$\left(\begin{array}{cc|c} 1 & 2 & 3 \\ 0 & 4 & -2 \\ 0 & 0 & 1 \end{array}\right)$$

or there is a unique solution as in the case where row operations yield an echelon form like

$$\left(\begin{array}{ccc|c} 1 & 2 & 2 & 3 \\ 0 & 4 & 3 & -2 \\ 0 & 0 & 4 & 1 \end{array}\right).$$

Also, sometimes there are free variables and no solution as in the following:

$$\left(\begin{array}{ccc|c} 1 & 2 & 2 & 3 \\ 0 & 4 & 3 & -2 \\ 0 & 0 & 0 & 1 \end{array}\right).$$

There are a lot of cases to consider but it is not necessary to make a major production of this. Do row operations till you obtain a matrix in echelon form or reduced echelon form and determine whether there is a solution. If there is, see if there are free variables. In this case, there will be infinitely many solutions. Find them by assigning different parameters to the free variables and obtain the solution. If there are no free variables, then there will be a unique solution which is easily determined once the augmented matrix is in echelon

or row reduced echelon form. In every case, the process yields a straightforward way to describe the solutions to the linear system. As indicated above, you are probably less likely to become confused if you place the augmented matrix in row reduced echelon form rather than just echelon form.

In summary,

Definition 3.2.19 *A **system of linear equations** is a list of equations,*

$$\begin{array}{c} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{array}$$

where a_{ij} are numbers, and b_j is a number. The above is a system of m equations in the n variables, $x_1, x_2 \cdots, x_n$. Nothing is said about the relative size of m and n. Written more simply in terms of summation notation, the above can be written in the form

$$\sum_{j=1}^{n} a_{ij}x_j = f_j,\ i = 1, 2, 3, \cdots, m$$

It is desired to find $(x_1, \cdots, x_n)$ solving each of the equations listed.

As illustrated above, such a system of linear equations may have a unique solution, no solution, or infinitely many solutions and these are the only three cases which can occur for any linear system. Furthermore, you do exactly the same things to solve any linear system. You write the augmented matrix and do row operations until you get a simpler system in which it is possible to see the solution, usually obtaining a matrix in echelon or reduced echelon form. All is based on the observation that the row operations do not change the solution set. You can have more equations than variables, fewer equations than variables, etc. It doesn't matter. You always set up the augmented matrix and go to work on it.

Definition 3.2.20 *A system of linear equations is called **consistent** if there exists a solution. It is called **inconsistent** if there is no solution.*

These are reasonable words to describe the situations of having or not having a solution. If you think of each equation as a condition which must be satisfied by the variables, consistent would mean there is some choice of variables which can satisfy all the conditions. Inconsistent would mean there is no choice of the variables which can satisfy each of the conditions.

Matrices

4.0.3 Outcomes

A. Perform the basic matrix operations of matrix addition, scalar multiplication, transposition and matrix multiplication. Identify when these operations are not defined. Represent the basic operations in terms of double subscript notation.

B. Recall and prove algebraic properties for matrix addition, scalar multiplication, transposition, and matrix multiplication. Apply these properties to manipulate an algebraic expression involving matrices.

C. Recall the cancelation laws for matrix multiplication. Demonstrate when cancelation laws do not apply.

D. Evaluate the inverse of a matrix using row operations.

E. Solve a linear system using matrix algebra.

F. Recall and prove identities involving matrix inverses.

4.1 Matrix Arithmetic

4.1.1 Addition And Scalar Multiplication Of Matrices

You have now solved systems of equations by writing them in terms of an augmented matrix and then doing row operations on this augmented matrix. It turns out such rectangular arrays of numbers are important from many other different points of view. Numbers are also called **scalars**. In these notes numbers will always be either real or complex numbers. I will refer to the set of numbers as $\mathbb{F}$ sometimes when it is not important to worry about whether the number is real or complex. Thus $\mathbb{F}$ can be either the real numbers, $\mathbb{R}$ or the complex numbers, $\mathbb{C}$.

A **matrix** is a rectangular array of numbers. Several of them are referred to as **matrices**. For example, here is a matrix.

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 2 & 8 & 7 \\ 6 & -9 & 1 & 2 \end{pmatrix}$$

The size or dimension of a matrix is defined as $m \times n$ where m is the number of rows and n is the number of columns. The above matrix is a 3×4 matrix because there are three rows and four columns. The first row is $(1\ 2\ 3\ 4)$, the second row is $(5\ 2\ 8\ 7)$ and so forth. The

first column is $\begin{pmatrix} 1 \\ 5 \\ 6 \end{pmatrix}$. When specifying the size of a matrix, you always list the number of rows before the number of columns. Also, you can remember the columns are like columns in a Greek temple. They stand upright while the rows just lay there like rows made by a tractor in a plowed field. Elements of the matrix are identified according to position in the matrix. For example, 8 is in position $2,3$ because it is in the second row and the third column. You might remember that you always list the rows before the columns by using the phrase **Row**man **C**atholic. The symbol, (a_{ij}) refers to a matrix. The entry in the i^{th} row and the j^{th} column of this matrix is denoted by a_{ij}. Using this notation on the above matrix, $a_{23} = 8, a_{32} = -9, a_{12} = 2$, etc.

There are various operations which are done on matrices. Matrices can be added multiplied by a scalar, and multiplied by other matrices. To illustrate scalar multiplication, consider the following example in which a matrix is being multiplied by the scalar, 3.

$$3\begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 2 & 8 & 7 \\ 6 & -9 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 6 & 9 & 12 \\ 15 & 6 & 24 & 21 \\ 18 & -27 & 3 & 6 \end{pmatrix}.$$

The new matrix is obtained by multiplying every entry of the original matrix by the given scalar. If A is an $m \times n$ matrix, $-A$ is defined to equal $(-1)\,A$.

Two matrices must be the same size to be added. The sum of two matrices is a matrix which is obtained by adding the corresponding entries. Thus

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 2 \end{pmatrix} + \begin{pmatrix} -1 & 4 \\ 2 & 8 \\ 6 & -4 \end{pmatrix} = \begin{pmatrix} 0 & 6 \\ 5 & 12 \\ 11 & -2 \end{pmatrix}.$$

Two matrices are equal exactly when they are the same size and the corresponding entries are identical. Thus

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

because they are different sizes. As noted above, you write (c_{ij}) for the matrix C whose ij^{th} entry is c_{ij}. In doing arithmetic with matrices you must define what happens in terms of the c_{ij} sometimes called the **entries** of the matrix or the **components** of the matrix.

The above discussion stated for general matrices is given in the following definition.

Definition 4.1.1 *(Scalar Multiplication) If $A = (a_{ij})$ and k is a scalar, then $kA = (ka_{ij})$.*

Example 4.1.2 $7\begin{pmatrix} 2 & 0 \\ 1 & -4 \end{pmatrix} = \begin{pmatrix} 14 & 0 \\ 7 & -28 \end{pmatrix}$.

Definition 4.1.3 *(Addition) If $A = (a_{ij})$ and $B = (b_{ij})$ are two $m \times n$ matrices. Then $A + B = C$ where*

$$C = (c_{ij})$$

for $c_{ij} = a_{ij} + b_{ij}$.

Example 4.1.4

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 4 \end{pmatrix} + \begin{pmatrix} 5 & 2 & 3 \\ -6 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 4 & 6 \\ -5 & 2 & 5 \end{pmatrix}$$

To save on notation, we will often use A_{ij} to refer to the ij^{th} entry of the matrix, A.

Definition 4.1.5 *(The zero matrix) The $m \times n$ zero matrix is the $m \times n$ matrix having every entry equal to zero. It is denoted by* 0.

Example 4.1.6 *The 2×3 zero matrix is* $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

Note there are 2×3 zero matrices, 3×4 zero matrices, etc. In fact there is a zero matrix for every size.

Definition 4.1.7 *(Equality of matrices) Let A and B be two matrices. Then $A = B$ means that the two matrices are of the same size and for $A = (a_{ij})$ and $B = (b_{ij})$, $a_{ij} = b_{ij}$ for all $1 \leq i \leq m$ and $1 \leq j \leq n$.*

The following properties of matrices can be easily verified. You should do so.

- Commutative Law Of Addition.

$$A + B = B + A, \tag{4.1}$$

- Associative Law for Addition.

$$(A + B) + C = A + (B + C), \tag{4.2}$$

- Existence of an Additive Identity

$$A + 0 = A, \tag{4.3}$$

- Existence of an Additive Inverse

$$A + (-A) = 0, \tag{4.4}$$

Also for α, β scalars, the following additional properties hold.

- Distributive law over Matrix Addition.

$$\alpha (A + B) = \alpha A + \alpha B, \tag{4.5}$$

- Distributive law over Scalar Addition

$$(\alpha + \beta) A = \alpha A + \beta A, \tag{4.6}$$

- Associative law for Scalar Multiplication

$$\alpha (\beta A) = \alpha\beta (A), \tag{4.7}$$

- Rule for Multiplication by 1.

$$1A = A. \tag{4.8}$$

As an example, consider the Commutative Law of Addition. Let $A + B = C$ and $B + A = D$. Why is $D = C$?

$$C_{ij} = A_{ij} + B_{ij} = B_{ij} + A_{ij} = D_{ij}.$$

Therefore, $C = D$ because the ij^{th} entries are the same. Note that the conclusion follows from the commutative law of addition of numbers.

4.1.2 Multiplication Of Matrices

Definition 4.1.8 *Matrices which are $n\times 1$ or $1\times n$ are called **vectors** and are often denoted by a bold letter. Thus the $n\times 1$ matrix*

$$\mathbf{x}=\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

*is also called a **column vector.** The $1\times n$ matrix*

$$(x_1\cdots x_n)$$

*is called a **row vector**.*

Although the following description of matrix multiplication may seem strange, it is in fact the most important and useful of the matrix operations. To begin with consider the case where a matrix is multiplied by a column vector. First consider a special case.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}\begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix} =?$$

One way to remember this is as follows. Slide the vector, placing it on top the two rows as shown and then do the indicated operation.

$$\begin{pmatrix} \mathbf{7} & \mathbf{8} & \mathbf{9} \\ 1 & 2 & 3 \\ \mathbf{7} & \mathbf{8} & \mathbf{9} \\ 4 & 5 & 6 \end{pmatrix}\rightarrow\begin{pmatrix} 7\times 1+8\times 2+9\times 3 \\ 7\times 4+8\times 5+9\times 6 \end{pmatrix}=\begin{pmatrix} 50 \\ 122 \end{pmatrix}.$$

multiply the numbers on the top by the numbers on the bottom and add them up to get a single number for each row of the matrix as shown above.

In more general terms,

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}=\begin{pmatrix} a_{11}x_1+a_{12}x_2+a_{13}x_3 \\ a_{21}x_1+a_{22}x_2+a_{23}x_3 \end{pmatrix}.$$

Another way to think of this is

$$x_1\begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}+x_2\begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}+x_3\begin{pmatrix} a_{13} \\ a_{23} \end{pmatrix}$$

Thus you take x_1 times the first column, add to x_2 times the second column, and finally x_3 times the third column. In general, here is the definition of how to multiply an $(m\times n)$ matrix times a $(n\times 1)$ matrix.

Definition 4.1.9 *Let $A=A_{ij}$ be an $m\times n$ matrix and let $\mathbf{v}$ be an $n\times 1$ matrix,*

$$\mathbf{v}=\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

Then $A\mathbf{v}$ is an $m\times 1$ matrix and the i^{th} component of this matrix is

$$(A\mathbf{v})_i=A_{i1}v_1+A_{i2}v_2+\cdots+A_{in}v_n=\sum_{j=1}^{n}A_{ij}v_j.$$

Thus

$$A\mathbf{v} = \begin{pmatrix} \sum_{j=1}^{n} A_{1j}v_j \\ \vdots \\ \sum_{j=1}^{n} A_{mj}v_j \end{pmatrix}. \tag{4.9}$$

In other words, if

$$A = (\mathbf{a}_1, \cdots, \mathbf{a}_n)$$

where the $\mathbf{a}_k$ *are the columns,*

$$A\mathbf{v} = \sum_{k=1}^{n} v_k \mathbf{a}_k$$

This follows from 4.9 and the observation that the j^{th} *column of* A *is*

$$\begin{pmatrix} A_{1j} \\ A_{2j} \\ \vdots \\ A_{mj} \end{pmatrix}$$

so 4.9 reduces to

$$v_1 \begin{pmatrix} A_{11} \\ A_{21} \\ \vdots \\ A_{m1} \end{pmatrix} + v_2 \begin{pmatrix} A_{12} \\ A_{22} \\ \vdots \\ A_{m2} \end{pmatrix} + \cdots + v_n \begin{pmatrix} A_{1n} \\ A_{2n} \\ \vdots \\ A_{mn} \end{pmatrix}$$

Note also that multiplication by an $m \times n$ *matrix takes an* $n \times 1$ *matrix, and produces an* $m \times 1$ *matrix.*

Here is another example.

Example 4.1.10 *Compute*

$$\begin{pmatrix} 1 & 2 & 1 & 3 \\ 0 & 2 & 1 & -2 \\ 2 & 1 & 4 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix}.$$

First of all this is of the form $(3 \times 4)(4 \times 1)$ and so the result should be a (3×1). Note how the inside numbers cancel. To get the element in the second row and first and only column, compute

$$\begin{aligned} \sum_{k=1}^{4} a_{2k}v_k &= a_{21}v_1 + a_{22}v_2 + a_{23}v_3 + a_{24}v_4 \\ &= 0 \times 1 + 2 \times 2 + 1 \times 0 + (-2) \times 1 = 2. \end{aligned}$$

You should do the rest of the problem and verify

$$\begin{pmatrix} 1 & 2 & 1 & 3 \\ 0 & 2 & 1 & -2 \\ 2 & 1 & 4 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 8 \\ 2 \\ 5 \end{pmatrix}.$$

The next task is to multiply an $m \times n$ matrix times an $n \times p$ matrix. Before doing so, the following may be helpful.

For A and B matrices, in order to form the product, AB the number of columns of A must equal the number of rows of B.

$$\begin{array}{c} \text{these must match!} \\ (m \times \quad \widehat{n)\,(n} \times p \quad) = m \times p \end{array}$$

Note the two outside numbers give the size of the product. Remember:

If the two middle numbers don't match, you can't multiply the matrices!

Definition 4.1.11 *When the number of columns of A equals the number of rows of B the two matrices are said to be* ***conformable*** *and the product, AB is obtained as follows. Let A be an $m \times n$ matrix and let B be an $n \times p$ matrix. Then B is of the form*

$$B = (\mathbf{b}_1, \cdots, \mathbf{b}_p)$$

where $\mathbf{b}_k$ is an $n \times 1$ matrix or column vector. Then the $m \times p$ matrix, AB is defined as follows:

$$AB \equiv (A\mathbf{b}_1, \cdots, A\mathbf{b}_p) \tag{4.10}$$

where $A\mathbf{b}_k$ is an $m \times 1$ matrix or column vector which gives the k^{th} column of AB.

Example 4.1.12 *Multiply the following.*

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 3 & 1 \\ -2 & 1 & 1 \end{pmatrix}$$

The first thing you need to check before doing anything else is whether it is possible to do the multiplication. The first matrix is a 2×3 and the second matrix is a 3×3. Therefore, is it possible to multiply these matrices. According to the above discussion it should be a 2×3 matrix of the form

$$\left(\overbrace{\begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -2 \end{pmatrix}}^{\text{First column}}, \overbrace{\begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}}^{\text{Second column}}, \overbrace{\begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}}^{\text{Third column}} \right)$$

You know how to multiply a matrix times a vector and so you do so to obtain each of the three columns. Thus

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 0 & 3 & 1 \\ -2 & 1 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 9 & 3 \\ -2 & 7 & 3 \end{pmatrix}.$$

Example 4.1.13 *Multiply the following.*

$$\begin{pmatrix} 1 & 2 & 0 \\ 0 & 3 & 1 \\ -2 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1 \end{pmatrix}$$

First check if it is possible. This is of the form $(3 \times 3)(2 \times 3)$. The inside numbers do not match and so you can't do this multiplication. This means that anything you write will be absolute nonsense because it is impossible to multiply these matrices in this order. Aren't they the same two matrices considered in the previous example? Yes they are. It is just that here they are in a different order. This shows something you must always remember about matrix multiplication.

Order Matters!

Matrix Multiplication Is Not Commutative!

This is very different than multiplication of numbers!

4.1.3 The ij^{th} Entry Of A Product

It is important to describe matrix multiplication in terms of entries of the matrices. What is the ij^{th} entry of AB? It would be the i^{th} entry of the j^{th} column of AB. Thus it would be the i^{th} entry of $A\mathbf{b}_j$. Now

$$\mathbf{b}_j = \begin{pmatrix} B_{1j} \\ \vdots \\ B_{nj} \end{pmatrix}$$

and from the above definition, the i^{th} entry is

$$\sum_{k=1}^{n} A_{ik} B_{kj}. \tag{4.11}$$

In terms of pictures of the matrix, you are doing

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1p} \\ B_{21} & B_{22} & \cdots & B_{2p} \\ \vdots & \vdots & & \vdots \\ B_{n1} & B_{n2} & \cdots & B_{np} \end{pmatrix}$$

Then as explained above, the j^{th} column is of the form

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix} \begin{pmatrix} B_{1j} \\ B_{2j} \\ \vdots \\ B_{nj} \end{pmatrix}$$

which is a $m \times 1$ matrix or column vector which equals

$$\begin{pmatrix} A_{11} \\ A_{21} \\ \vdots \\ A_{m1} \end{pmatrix} B_{1j} + \begin{pmatrix} A_{12} \\ A_{22} \\ \vdots \\ A_{m2} \end{pmatrix} B_{2j} + \cdots + \begin{pmatrix} A_{1n} \\ A_{2n} \\ \vdots \\ A_{mn} \end{pmatrix} B_{nj}.$$

The second entry of this $m \times 1$ matrix is

$$A_{21}B_{1j} + A_{22}B_{2j} + \cdots + A_{2n}B_{nj} = \sum_{k=1}^{m} A_{2k} B_{kj}.$$

Similarly, the i^{th} entry of this $m \times 1$ matrix is

$$A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{in}B_{nj} = \sum_{k=1}^{m} A_{ik}B_{kj}.$$

This shows the following definition for matrix multiplication in terms of the ij^{th} entries of the product coincides with Definition 4.1.11.

Definition 4.1.14 *Let $A = (A_{ij})$ be an $m \times n$ matrix and let $B = (B_{ij})$ be an $n \times p$ matrix. Then AB is an $m \times p$ matrix and*

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj}. \tag{4.12}$$

Another way to write this is

$$(AB)_{ij} = \begin{pmatrix} A_{i1} & A_{i2} & \cdots & A_{in} \end{pmatrix} \begin{pmatrix} B_{1j} \\ B_{2j} \\ \vdots \\ B_{nj} \end{pmatrix}$$

Note that to get $(AB)_{ij}$ you involve the i^{th} row of A and the j^{th} column of B.

Example 4.1.15 *Multiply if possible* $\begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 2 & 3 & 1 \\ 7 & 6 & 2 \end{pmatrix}$.

First check to see if this is possible. It is of the form $(3 \times 2)(2 \times 3)$ and since the inside numbers match, the two matrices are conformable and it is possible to do the multiplication. The result should be a 3×3 matrix. The answer is of the form

$$\left(\begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 2 \\ 7 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 3 \\ 6 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right)$$

where the commas separate the columns in the resulting product. Thus the above product equals

$$\begin{pmatrix} 16 & 15 & 5 \\ 13 & 15 & 5 \\ 46 & 42 & 14 \end{pmatrix},$$

a 3×3 matrix as desired. In terms of the ij^{th} entries and the above definition, the entry in the third row and second column of the product should equal

$$\begin{aligned} \sum_{j} a_{3k}b_{k2} &= a_{31}b_{12} + a_{32}b_{22} \\ &= 2 \times 3 + 6 \times 6 = 42. \end{aligned}$$

You should try a few more such examples to verify the above definition in terms of the ij^{th} entries works for other entries.

Example 4.1.16 *Multiply if possible* $\begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix} \begin{pmatrix} 2 & 3 & 1 \\ 7 & 6 & 2 \\ 0 & 0 & 0 \end{pmatrix}$.

This is not possible because it is of the form $(3\times 2)(3\times 3)$ and the middle numbers don't match. In other words the two matrices are not conformable in the indicated order.

Example 4.1.17 *Multiply if possible* $\begin{pmatrix} 2 & 3 & 1 \\ 7 & 6 & 2 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 2 & 6 \end{pmatrix}$.

This is possible because in this case it is of the form $(3\times 3)(3\times 2)$ and the middle numbers do match so the matrices are conformable. When the multiplication is done it equals

$$\begin{pmatrix} 13 & 13 \\ 29 & 32 \\ 0 & 0 \end{pmatrix}.$$

Check this and be sure you come up with the same answer.

Example 4.1.18 *Multiply if possible* $\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & 2 & 1 & 0 \end{pmatrix}$.

In this case you are trying to do $(3\times 1)(1\times 4)$. The inside numbers match so you can do it. Verify

$$\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & 2 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 1 & 0 \\ 2 & 4 & 2 & 0 \\ 1 & 2 & 1 & 0 \end{pmatrix}$$

4.1.4 Properties Of Matrix Multiplication

As pointed out above, sometimes it is possible to multiply matrices in one order but not in the other order. What if it makes sense to multiply them in either order? Will the two products be equal then?

Example 4.1.19 *Compare* $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ *and* $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$.

The first product is

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix}.$$

The second product is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix}.$$

You see these are not equal. Again you cannot conclude that $AB = BA$ for matrix multiplication even when multiplication is defined in both orders. However, there are some properties which do hold.

Proposition 4.1.20 *If all multiplications and additions make sense, the following hold for matrices,* A, B, C *and* a, b *scalars.*

$$A(aB + bC) = a(AB) + b(AC) \tag{4.13}$$

$$(B + C)A = BA + CA \tag{4.14}$$

$$A(BC) = (AB)C \tag{4.15}$$

Proof: Using Definition 4.1.14,

$$\begin{aligned}
\left(A\left(aB+bC\right)\right)_{ij} &= \sum_k A_{ik}\left(aB+bC\right)_{kj} \\
&= \sum_k A_{ik}\left(aB_{kj}+bC_{kj}\right) \\
&= a\sum_k A_{ik}B_{kj} + b\sum_k A_{ik}C_{kj} \\
&= a\left(AB\right)_{ij} + b\left(AC\right)_{ij} \\
&= \left(a\left(AB\right)+b\left(AC\right)\right)_{ij}.
\end{aligned}$$

Thus $A\left(B+C\right) = AB+AC$ as claimed. Formula 4.14 is entirely similar.

Formula 4.15 is the associative law of multiplication. Using Definition 4.1.14,

$$\begin{aligned}
\left(A\left(BC\right)\right)_{ij} &= \sum_k A_{ik}\left(BC\right)_{kj} \\
&= \sum_k A_{ik}\sum_l B_{kl}C_{lj} \\
&= \sum_l \left(AB\right)_{il}C_{lj} \\
&= \left(\left(AB\right)C\right)_{ij}.
\end{aligned}$$

This proves 4.15.

4.1.5 The Transpose

Another important operation on matrices is that of taking the **transpose**. The following example shows what is meant by this operation, denoted by placing a T as an exponent on the matrix.

$$\begin{pmatrix} 1 & 4 \\ 3 & 1 \\ 2 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 & 2 \\ 4 & 1 & 6 \end{pmatrix}$$

What happened? The first column became the first row and the second column became the second row. Thus the 3×2 matrix became a 2×3 matrix. The number 3 was in the second row and the first column and it ended up in the first row and second column. Here is the definition.

Definition 4.1.21 *Let A be an $m\times n$ matrix. Then A^T denotes the $n\times m$ matrix which is defined as follows.*

$$\left(A^T\right)_{ij} = A_{ji}$$

Example 4.1.22

$$\begin{pmatrix} 1 & 2 & -6 \\ 3 & 5 & 4 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 \\ 2 & 5 \\ -6 & 4 \end{pmatrix}.$$

The transpose of a matrix has the following important properties.

Lemma 4.1.23 *Let A be an $m\times n$ matrix and let B be a $n\times p$ matrix. Then*

$$\left(AB\right)^T = B^TA^T \tag{4.16}$$

and if α and β are scalars,

$$(\alpha A + \beta B)^T = \alpha A^T + \beta B^T \tag{4.17}$$

Proof: From the definition,

$$\begin{aligned}\left((AB)^T\right)_{ij} &= (AB)_{ji} \\ &= \sum_k A_{jk}B_{ki} \\ &= \sum_k \left(B^T\right)_{ik}\left(A^T\right)_{kj} \\ &= \left(B^T A^T\right)_{ij}\end{aligned}$$

The proof of Formula 4.17 is left as an exercise and this proves the lemma.

Definition 4.1.24 *An $n \times n$ matrix, A is said to be **symmetric** if $A = A^T$. It is said to be **skew symmetric** if $A = -A^T$.*

Example 4.1.25 *Let*

$$A = \begin{pmatrix} 2 & 1 & 3 \\ 1 & 5 & -3 \\ 3 & -3 & 7 \end{pmatrix}.$$

Then A is symmetric.

Example 4.1.26 *Let*

$$A = \begin{pmatrix} 0 & 1 & 3 \\ -1 & 0 & 2 \\ -3 & -2 & 0 \end{pmatrix}$$

Then A is skew symmetric.

4.1.6 The Identity And Inverses

There is a special matrix called I and referred to as the identity matrix. It is always a square matrix, meaning the number of rows equals the number of columns and it has the property that there are ones down the main diagonal and zeroes elsewhere. Here are some identity matrices of various sizes.

$$(1), \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The first is the 1×1 identity matrix, the second is the 2×2 identity matrix, the third is the 3×3 identity matrix, and the fourth is the 4×4 identity matrix. By extension, you can likely see what the $n \times n$ identity matrix would be. It is so important that there is a special symbol to denote the ij^{th} entry of the identity matrix

$$I_{ij} = \delta_{ij}$$

where δ_{ij} is the **Kroneker symbol** defined by

$$\delta_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

It is called the **identity matrix** because it is a **multiplicative identity** in the following sense.

Lemma 4.1.27 *Suppose A is an $m \times n$ matrix and I_n is the $n \times n$ identity matrix. Then $AI_n = A$. If I_m is the $m \times m$ identity matrix, it also follows that $I_m A = A$.*

Proof:

$$\begin{aligned}(AI_n)_{ij} &= \sum_k A_{ik}\delta_{kj} \\ &= A_{ij}\end{aligned}$$

and so $AI_n = A$. The other case is left as an exercise for you.

Definition 4.1.28 *An $n \times n$ matrix, A has an **inverse**, A^{-1} if and only if $AA^{-1} = A^{-1}A = I$. Such a matrix is called **invertible**.*

It is very important to observe that the inverse of a matrix, if it exists, is unique. Another way to think of this is that if it acts like the inverse, then it is the inverse.

Theorem 4.1.29 *Suppose A^{-1} exists and $AB = BA = I$. Then $B = A^{-1}$.*

Proof:

$$A^{-1} = A^{-1}I = A^{-1}(AB) = \left(A^{-1}A\right)B = IB = B.$$

Unlike ordinary multiplication of numbers, it can happen that $A \neq 0$ but A may fail to have an inverse. This is illustrated in the following example.

Example 4.1.30 *Let $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. Does A have an inverse?*

One might think A would have an inverse because it does not equal zero. However,

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and if A^{-1} existed, this could not happen because you could write

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = A^{-1}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) = A^{-1}\left(A\begin{pmatrix} -1 \\ 1 \end{pmatrix}\right) =$$

$$= \left(A^{-1}A\right)\begin{pmatrix} -1 \\ 1 \end{pmatrix} = I\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix},$$

a contradiction. Thus the answer is that A does not have an inverse.

Example 4.1.31 *Let $A = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. Show $\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$ is the inverse of A.*

To check this, multiply

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

showing that this matrix is indeed the inverse of A.

4.1.7 Finding The Inverse Of A Matrix

In the last example, how would you find A^{-1}? You wish to find a matrix, $\begin{pmatrix} x & z \\ y & w \end{pmatrix}$ such that

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x & z \\ y & w \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

This requires the solution of the systems of equations,

$$x + y = 1, x + 2y = 0$$

and

$$z + w = 0, z + 2w = 1.$$

Writing the augmented matrix for these two systems gives

$$\left(\begin{array}{cc|c} 1 & 1 & 1 \\ 1 & 2 & 0 \end{array} \right) \tag{4.18}$$

for the first system and

$$\left(\begin{array}{cc|c} 1 & 1 & 0 \\ 1 & 2 & 1 \end{array} \right) \tag{4.19}$$

for the second. Lets solve the first system. Take (-1) times the first row and add to the second to get

$$\left(\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & 1 & -1 \end{array} \right)$$

Now take (-1) times the second row and add to the first to get

$$\left(\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & -1 \end{array} \right).$$

Putting in the variables, this says $x = 2$ and $y = -1$.

Now solve the second system, 4.19 to find z and w. Take (-1) times the first row and add to the second to get

$$\left(\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 1 & 1 \end{array} \right).$$

Now take (-1) times the second row and add to the first to get

$$\left(\begin{array}{cc|c} 1 & 0 & -1 \\ 0 & 1 & 1 \end{array} \right).$$

Putting in the variables, this says $z = -1$ and $w = 1$. Therefore, the inverse is

$$\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}.$$

Didn't the above seem rather repetitive? Note that exactly the same row operations were used in both systems. In each case, the end result was something of the form $(I|\mathbf{v})$ where I is the identity and $\mathbf{v}$ gave a column of the inverse. In the above, $\begin{pmatrix} x \\ y \end{pmatrix}$, the first column of the inverse was obtained first and then the second column $\begin{pmatrix} z \\ w \end{pmatrix}$.

To simplify this procedure, you could have written

$$\left(\begin{array}{cc|cc} 1 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{array}\right)$$

and row reduced till you obtained

$$\left(\begin{array}{cc|cc} 1 & 0 & 2 & -1 \\ 0 & 1 & -1 & 1 \end{array}\right)$$

and read off the inverse as the 2×2 matrix on the right side.

This is the reason for the following simple procedure for finding the inverse of a matrix. This procedure is called the **Gauss-Jordan procedure**.

Procedure 4.1.32 *Suppose A is an $n \times n$ matrix. To find A^{-1} if it exists, form the augmented $n \times 2n$ matrix,*

$$(A|I)$$

and then, if possible do row operations until you obtain an $n \times 2n$ matrix of the form

$$(I|B)\,. \tag{4.20}$$

When this has been done, $B = A^{-1}$. If it is impossible to row reduce to a matrix of the form $(I|B)$, then A has no inverse.

Example 4.1.33 *Let* $A = \left(\begin{array}{ccc} 1 & 2 & 2 \\ 1 & 0 & 2 \\ 3 & 1 & -1 \end{array}\right)$. *Find A^{-1} if it exists.*

Set up the augmented matrix, $(A|I)$

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 2 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 & 1 & 0 \\ 3 & 1 & -1 & 0 & 0 & 1 \end{array}\right)$$

Next take (-1) times the first row and add to the second followed by (-3) times the first row added to the last. This yields

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & -2 & 0 & -1 & 1 & 0 \\ 0 & -5 & -7 & -3 & 0 & 1 \end{array}\right).$$

Then take 5 times the second row and add to -2 times the last row.

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & -10 & 0 & -5 & 5 & 0 \\ 0 & 0 & 14 & 1 & 5 & -2 \end{array}\right)$$

Next take the last row and add to (-7) times the top row. This yields

$$\left(\begin{array}{ccc|ccc} -7 & -14 & 0 & -6 & 5 & -2 \\ 0 & -10 & 0 & -5 & 5 & 0 \\ 0 & 0 & 14 & 1 & 5 & -2 \end{array}\right).$$

Now take $(-7/5)$ times the second row and add to the top.

$$\left(\begin{array}{ccc|ccc} -7 & 0 & 0 & 1 & -2 & -2 \\ 0 & -10 & 0 & -5 & 5 & 0 \\ 0 & 0 & 14 & 1 & 5 & -2 \end{array}\right).$$

Finally divide the top row by -7, the second row by -10 and the bottom row by 14 which yields

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{1}{7} & \frac{2}{7} & \frac{2}{7} \\ 0 & 1 & 0 & \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 1 & \frac{1}{14} & \frac{5}{14} & -\frac{1}{7} \end{array}\right).$$

Therefore, the inverse is

$$\left(\begin{array}{ccc} -\frac{1}{7} & \frac{2}{7} & \frac{2}{7} \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ \frac{1}{14} & \frac{5}{14} & -\frac{1}{7} \end{array}\right)$$

Example 4.1.34 *Let* $A = \begin{pmatrix} 1 & 2 & 2 \\ 1 & 0 & 2 \\ 2 & 2 & 4 \end{pmatrix}$. *Find* A^{-1} *if it exists.*

Write the augmented matrix, $(A|I)$

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 2 & 1 & 0 & 0 \\ 1 & 0 & 2 & 0 & 1 & 0 \\ 2 & 2 & 4 & 0 & 0 & 1 \end{array}\right)$$

and proceed to do row operations attempting to obtain $\left(I|A^{-1}\right)$. Take (-1) times the top row and add to the second. Then take (-2) times the top row and add to the bottom.

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & -2 & 0 & -1 & 1 & 0 \\ 0 & -2 & 0 & -2 & 0 & 1 \end{array}\right)$$

Next add (-1) times the second row to the bottom row.

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 2 & 1 & 0 & 0 \\ 0 & -2 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 1 \end{array}\right)$$

At this point, you can see there will be no inverse because you have obtained a row of zeros in the left half of the augmented matrix, $(A|I)$. Thus there will be no way to obtain I on the left.

Example 4.1.35 *Let* $A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}$. *Find* A^{-1} *if it exists.*

Form the augmented matrix,

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & -1 & 1 & 0 & 1 & 0 \\ 1 & 1 & -1 & 0 & 0 & 1 \end{array}\right).$$

Now do row operations until the $n \times n$ matrix on the left becomes the identity matrix. This yields after some computations,

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 1 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 1 & -\frac{1}{2} & -\frac{1}{2} \end{array}\right)$$

and so the inverse of A is the matrix on the right,

$$\left(\begin{array}{ccc} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & -1 & 0 \\ 1 & -\frac{1}{2} & -\frac{1}{2} \end{array}\right).$$

Checking the answer is easy. Just multiply the matrices and see if it works.

$$\left(\begin{array}{ccc} 1 & 0 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{array}\right)\left(\begin{array}{ccc} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & -1 & 0 \\ 1 & -\frac{1}{2} & -\frac{1}{2} \end{array}\right) = \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right).$$

Always check your answer because if you are like some of us, you will usually have made a mistake.

Example 4.1.36 *In this example, it is shown how to use the inverse of a matrix to find the solution to a system of equations. Consider the following system of equations. Use the inverse of a suitable matrix to give the solutions to this system.*

$$\left(\begin{array}{c} x + z = 1 \\ x - y + z = 3 \\ x + y - z = 2 \end{array}\right).$$

The system of equations can be written in terms of matrices as

$$\left(\begin{array}{ccc} 1 & 0 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{array}\right)\left(\begin{array}{c} x \\ y \\ z \end{array}\right) = \left(\begin{array}{c} 1 \\ 3 \\ 2 \end{array}\right). \tag{4.21}$$

More simply, this is of the form $A\mathbf{x} = \mathbf{b}$. Suppose you find the inverse of the matrix, A^{-1}. Then you could multiply both sides of this equation by A^{-1} to obtain

$$\mathbf{x} = \left(A^{-1}A\right)\mathbf{x} = A^{-1}\left(A\mathbf{x}\right) = A^{-1}\mathbf{b}.$$

This gives the solution as $\mathbf{x} = A^{-1}\mathbf{b}$. Note that once you have found the inverse, you can easily get the solution for different right hand sides without any effort. It is always just $A^{-1}\mathbf{b}$. In the given example, the inverse of the matrix is

$$\left(\begin{array}{ccc} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & -1 & 0 \\ 1 & -\frac{1}{2} & -\frac{1}{2} \end{array}\right)$$

This was shown in Example 4.1.35. Therefore, from what was just explained the solution to the given system is

$$\left(\begin{array}{c} x \\ y \\ z \end{array}\right) = \left(\begin{array}{ccc} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & -1 & 0 \\ 1 & -\frac{1}{2} & -\frac{1}{2} \end{array}\right)\left(\begin{array}{c} 1 \\ 3 \\ 2 \end{array}\right) = \left(\begin{array}{c} \frac{5}{2} \\ -2 \\ -\frac{3}{2} \end{array}\right).$$

What if the right side of 4.21 had been

$$\begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}?$$

What would be the solution to

$$\begin{pmatrix} 1 & 0 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix}?$$

By the above discussion, it is just

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & -1 & 0 \\ 1 & -\frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ -1 \\ -2 \end{pmatrix}.$$

This illustrates why once you have found the inverse of a given matrix, you can use it to solve many different systems easily.

Vector Products

5.0.8 Outcomes

1. Evaluate a dot product from the angle formula or the coordinate formula.
2. Interpret the dot product geometrically.
3. Evaluate the following using the dot product:
   (a) the angle between two vectors
   (b) the magnitude of a vector
   (c) the work done by a constant force on an object
4. Evaluate a cross product from the angle formula or the coordinate formula.
5. Interpret the cross product geometrically.
6. Evaluate the following using the cross product:
   (a) the area of a parallelogram
   (b) the area of a triangle
   (c) physical quantities such as the torque and angular velocity.
7. Find the volume of a parallelepiped using the box product.
8. Recall, apply and derive the algebraic properties of the dot and cross products.

5.1 The Dot Product

There are two ways of multiplying vectors which are of great importance in applications. The first of these is called the **dot product**, also called the **scalar product** and sometimes the **inner product**.

Definition 5.1.1 *Let* $\mathbf{a}, \mathbf{b}$ *be two vectors in* $\mathbb{R}^n$ *define* $\mathbf{a} \cdot \mathbf{b}$ *as*

$$\mathbf{a} \cdot \mathbf{b} \equiv \sum_{k=1}^{n} a_k b_k.$$

With this definition, there are several important properties satisfied by the dot product. In the statement of these properties, α and β will denote scalars and $\mathbf{a}, \mathbf{b}, \mathbf{c}$ will denote vectors.

Proposition 5.1.2 *The dot product satisfies the following properties.*

$$\mathbf{a}\cdot\mathbf{b}=\mathbf{b}\cdot\mathbf{a} \tag{5.1}$$

$$\mathbf{a}\cdot\mathbf{a}\geq 0 \text{ and equals zero if and only if } \mathbf{a}=\mathbf{0} \tag{5.2}$$

$$(\alpha\mathbf{a}+\beta\mathbf{b})\cdot\mathbf{c}=\alpha(\mathbf{a}\cdot\mathbf{c})+\beta(\mathbf{b}\cdot\mathbf{c}) \tag{5.3}$$

$$\mathbf{c}\cdot(\alpha\mathbf{a}+\beta\mathbf{b})=\alpha(\mathbf{c}\cdot\mathbf{a})+\beta(\mathbf{c}\cdot\mathbf{b}) \tag{5.4}$$

$$|\mathbf{a}|^2=\mathbf{a}\cdot\mathbf{a} \tag{5.5}$$

You should verify these properties. Also be sure you understand that 5.4 follows from the first three and is therefore redundant. It is listed here for the sake of convenience.

Example 5.1.3 *Find* $(1,2,0,-1)\cdot(0,1,2,3)$.

This equals $0+2+0+-3=-1$.

Example 5.1.4 *Find the magnitude of* $\mathbf{a}=(2,1,4,2)$. *That is, find* $|\mathbf{a}|$.

This is $\sqrt{(2,1,4,2)\cdot(2,1,4,2)}=5$.

The dot product satisfies a fundamental inequality known as the **Cauchy Schwarz inequality.**

Theorem 5.1.5 *The dot product satisfies the inequality*

$$|\mathbf{a}\cdot\mathbf{b}|\leq|\mathbf{a}||\mathbf{b}|. \tag{5.6}$$

Furthermore equality is obtained if and only if one of $\mathbf{a}$ *or* $\mathbf{b}$ *is a scalar multiple of the other.*

Proof: First note that if $\mathbf{b}=\mathbf{0}$ both sides of 5.6 equal zero and so the inequality holds in this case. Therefore, it will be assumed in what follows that $\mathbf{b}\neq\mathbf{0}$.

Define a function of $t\in\mathbb{R}$

$$f(t)=(\mathbf{a}+t\mathbf{b})\cdot(\mathbf{a}+t\mathbf{b}).$$

Then by 5.2, $f(t)\geq 0$ for all $t\in\mathbb{R}$. Also from 5.3,5.4,5.1, and 5.5

$$\begin{aligned}f(t)&=\mathbf{a}\cdot(\mathbf{a}+t\mathbf{b})+t\mathbf{b}\cdot(\mathbf{a}+t\mathbf{b})\\&=\mathbf{a}\cdot\mathbf{a}+t(\mathbf{a}\cdot\mathbf{b})+t\mathbf{b}\cdot\mathbf{a}+t^2\mathbf{b}\cdot\mathbf{b}\\&=|\mathbf{a}|^2+2t(\mathbf{a}\cdot\mathbf{b})+|\mathbf{b}|^2t^2.\end{aligned}$$

Now this means the graph, $y=f(t)$ is a polynomial which opens up and either its vertex touches the t axis or else the entire graph is above the x axis. In the first case, there exists some t where $f(t)=0$ and this requires $\mathbf{a}+t\mathbf{b}=\mathbf{0}$ so one vector is a multiple of the other. Then clearly equality holds in 5.6. In the case where $\mathbf{b}$ is not a multiple of $\mathbf{a}$, it follows $f(t)>0$ for all t which says $f(t)$ has no real zeros and so from the quadratic formula,

$$(2(\mathbf{a}\cdot\mathbf{b}))^2-4|\mathbf{a}|^2|\mathbf{b}|^2<0$$

which is equivalent to $|(\mathbf{a}\cdot\mathbf{b})|<|\mathbf{a}||\mathbf{b}|$. This proves the theorem.

You should note that the entire argument was based only on the properties of the dot product listed in 5.1 - 5.5. This means that whenever something satisfies these properties, the Cauchy Schwartz inequality holds. There are many other instances of these properties besides vectors in $\mathbb{R}^n$.

The Cauchy Schwartz inequality allows a proof of the **triangle inequality** for distances in $\mathbb{R}^n$ in much the same way as the triangle inequality for the absolute value.

Theorem 5.1.6 *(Triangle inequality) For* $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$

$$|\mathbf{a}+\mathbf{b}| \leq |\mathbf{a}| + |\mathbf{b}| \tag{5.7}$$

and equality holds if and only if one of the vectors is a nonnegative scalar multiple of the other. Also

$$||\mathbf{a}| - |\mathbf{b}|| \leq |\mathbf{a}-\mathbf{b}| \tag{5.8}$$

Proof: By properties of the dot product and the Cauchy Schwartz inequality,

$$\begin{aligned}
|\mathbf{a}+\mathbf{b}|^2 &= (\mathbf{a}+\mathbf{b})\cdot(\mathbf{a}+\mathbf{b}) \\
&= (\mathbf{a}\cdot\mathbf{a}) + (\mathbf{a}\cdot\mathbf{b}) + (\mathbf{b}\cdot\mathbf{a}) + (\mathbf{b}\cdot\mathbf{b}) \\
&= |\mathbf{a}|^2 + 2(\mathbf{a}\cdot\mathbf{b}) + |\mathbf{b}|^2 \\
&\leq |\mathbf{a}|^2 + 2|\mathbf{a}\cdot\mathbf{b}| + |\mathbf{b}|^2 \\
&\leq |\mathbf{a}|^2 + 2|\mathbf{a}||\mathbf{b}| + |\mathbf{b}|^2 \\
&= (|\mathbf{a}| + |\mathbf{b}|)^2 .
\end{aligned}$$

Taking square roots of both sides you obtain 5.7.

It remains to consider when equality occurs. If either vector equals zero, then that vector equals zero times the other vector and the claim about when equality occurs is verified. Therefore, it can be assumed both vectors are nonzero. To get equality in the second inequality above, Theorem 5.1.5 implies one of the vectors must be a multiple of the other. Say $\mathbf{b} = \alpha\mathbf{a}$. If $\alpha < 0$ then equality cannot occur in the first inequality because in this case

$$(\mathbf{a}\cdot\mathbf{b}) = \alpha|\mathbf{a}|^2 < 0 < |\alpha||\mathbf{a}|^2 = |\mathbf{a}\cdot\mathbf{b}|$$

Therefore, $\alpha \geq 0$.

To get the other form of the triangle inequality,

$$\mathbf{a} = \mathbf{a}-\mathbf{b}+\mathbf{b}$$

so

$$\begin{aligned}
|\mathbf{a}| &= |\mathbf{a}-\mathbf{b}+\mathbf{b}| \\
&\leq |\mathbf{a}-\mathbf{b}| + |\mathbf{b}| .
\end{aligned}$$

Therefore,

$$|\mathbf{a}| - |\mathbf{b}| \leq |\mathbf{a}-\mathbf{b}| \tag{5.9}$$

Similarly,

$$|\mathbf{b}| - |\mathbf{a}| \leq |\mathbf{b}-\mathbf{a}| = |\mathbf{a}-\mathbf{b}| . \tag{5.10}$$

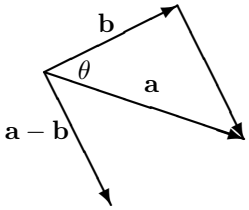
It follows from 5.9 and 5.10 that 5.8 holds. This is because $||\mathbf{a}| - |\mathbf{b}||$ equals the left side of either 5.9 or 5.10 and either way, $||\mathbf{a}| - |\mathbf{b}|| \leq |\mathbf{a}-\mathbf{b}|$. This proves the theorem.

5.2 The Geometric Significance Of The Dot Product

5.2.1 The Angle Between Two Vectors

Given two vectors, $\mathbf{a}$ and $\mathbf{b}$, the included angle is the angle between these two vectors which is less than or equal to 180 degrees. The dot product can be used to determine the included

angle between two vectors. To see how to do this, consider the following picture.



By the law of cosines,

$$|\mathbf{a}-\mathbf{b}|^2 = |\mathbf{a}|^2 + |\mathbf{b}|^2 - 2|\mathbf{a}||\mathbf{b}|\cos\theta.$$

Also from the properties of the dot product,

$$\begin{aligned}|\mathbf{a}-\mathbf{b}|^2 &= (\mathbf{a}-\mathbf{b})\cdot(\mathbf{a}-\mathbf{b})\\ &= |\mathbf{a}|^2 + |\mathbf{b}|^2 - 2\mathbf{a}\cdot\mathbf{b}\end{aligned}$$

and so comparing the above two formulas,

$$\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}||\mathbf{b}|\cos\theta. \tag{5.11}$$

In words, the dot product of two vectors equals the product of the magnitude of the two vectors multiplied by the cosine of the included angle. Note this gives a geometric description of the dot product which does not depend explicitly on the coordinates of the vectors.

Example 5.2.1 *Find the angle between the vectors* $2\mathbf{i}+\mathbf{j}-\mathbf{k}$ *and* $3\mathbf{i}+4\mathbf{j}+\mathbf{k}$.

The dot product of these two vectors equals $6+4-1=9$ and the norms are $\sqrt{4+1+1} = \sqrt{6}$ and $\sqrt{9+16+1} = \sqrt{26}$. Therefore, from 5.11 the cosine of the included angle equals

$$\cos\theta = \frac{9}{\sqrt{26}\sqrt{6}} = .720\,58$$

Now the cosine is known, the angle can be determines by solving the equation, $\cos\theta = .720\,58$. This will involve using a calculator or a table of trigonometric functions. The answer is $\theta = .766\,16$ radians or in terms of degrees, $\theta = .766\,16 \times \frac{360}{2\pi} = 43.898°$. Recall how this last computation is done. Set up a proportion, $\frac{x}{.766\,16} = \frac{360}{2\pi}$ because $360°$ corresponds to 2π radians. However, in calculus, you should get used to thinking in terms of radians and not degrees. This is because all the important calculus formulas are defined in terms of radians.

Example 5.2.2 *Let* $\mathbf{u},\mathbf{v}$ *be two vectors whose magnitudes are equal to 3 and 4 respectively and such that if they are placed in standard position with their tails at the origin, the angle between* $\mathbf{u}$ *and the positive x axis equals 30° and the angle between* $\mathbf{v}$ *and the positive x axis is -30°. Find* $\mathbf{u}\cdot\mathbf{v}$.

From the geometric description of the dot product in 5.11

$$\mathbf{u}\cdot\mathbf{v} = 3\times 4\times\cos(60°) = 3\times 4\times 1/2 = 6.$$

Observation 5.2.3 *Two vectors are said to be* ***perpendicular*** *if the included angle is* $\pi/2$ *radians (90°). You can tell if two nonzero vectors are perpendicular by simply taking their dot product. If the answer is zero, this means they are perpendicular because* $\cos\theta = 0$.

Example 5.2.4 *Determine whether the two vectors,* $2\mathbf{i}+\mathbf{j}-\mathbf{k}$ *and* $1\mathbf{i}+3\mathbf{j}+5\mathbf{k}$ *are perpendicular.*

When you take this dot product you get $2+3-5=0$ and so these two are indeed perpendicular.

Definition 5.2.5 *When two lines intersect, the angle between the two lines is the smaller of the two angles determined.*

Example 5.2.6 *Find the angle between the two lines,* $(1,2,0)+t(1,2,3)$ *and* $(0,4,-3)+t(-1,2,-3)$.

These two lines intersect, when $t=0$ in the first and $t=-1$ in the second. It is only a matter of finding the angle between the direction vectors. One angle determined is given by

$$\cos\theta=\frac{-6}{14}=\frac{-3}{7}. \tag{5.12}$$

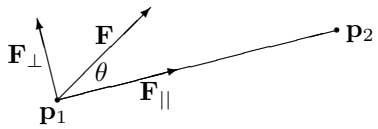
We don't want this angle because it is obtuse. The angle desired is the acute angle given by

$$\cos\theta=\frac{3}{7}.$$

It is obtained by using replacing one of the direction vectors with -1 times it.

5.2.2 Work And Projections

Our first application will be to the concept of work. The physical concept of work does not in any way correspond to the notion of work employed in ordinary conversation. For example, if you were to slide a 150 pound weight off a table which is three feet high and shuffle along the floor for 50 yards, sweating profusely and exerting all your strength to keep the weight from falling on your feet, keeping the height always three feet and then deposit this weight on another three foot high table, the physical concept of work would indicate that the force exerted by your arms did no work during this project even though the muscles in your hands and arms would likely be very tired. The reason for such an unusual definition is that even though your arms exerted considerable force on the weight, enough to keep it from falling, the direction of motion was at right angles to the force they exerted. The only part of a force which does work in the sense of physics is the component of the force in the direction of motion (This is made more precise below.). The work is defined to be the magnitude of the component of this force times the distance over which it acts in the case where this component of force points in the direction of motion and (-1) times the magnitude of this component times the distance in case the force tends to impede the motion. Thus the work done by a force on an object as the object moves from one point to another is a measure of the extent to which the force contributes to the motion. This is illustrated in the following picture in the case where the given force contributes to the motion.



In this picture the force, $\mathbf{F}$ is applied to an object which moves on the straight line from $\mathbf{p}_1$ to $\mathbf{p}_2$. There are two vectors shown, $\mathbf{F}_{||}$ and $\mathbf{F}_{\perp}$ and the picture is intended to indicate

that when you add these two vectors you get $\mathbf{F}$ while $\mathbf{F}_{||}$ acts in the direction of motion and $\mathbf{F}_{\perp}$ acts perpendicular to the direction of motion. Only $\mathbf{F}_{||}$ contributes to the work done by $\mathbf{F}$ on the object as it moves from $\mathbf{p}_1$ to $\mathbf{p}_2$. $\mathbf{F}_{||}$ is called the **component of the force** in the direction of motion. From trigonometry, you see the magnitude of $\mathbf{F}_{||}$ should equal $|\mathbf{F}||\cos\theta|$. Thus, since $\mathbf{F}_{||}$ points in the direction of the vector from $\mathbf{p}_1$ to $\mathbf{p}_2$, the total work done should equal

$$|\mathbf{F}|\left|\overrightarrow{\mathbf{p}_1\mathbf{p}_2}\right|\cos\theta = |\mathbf{F}||\mathbf{p}_2-\mathbf{p}_1|\cos\theta$$

If the included angle had been obtuse, then the work done by the force, $\mathbf{F}$ on the object would have been negative because in this case, the force tends to impede the motion from $\mathbf{p}_1$ to $\mathbf{p}_2$ but in this case, $\cos\theta$ would also be negative and so it is still the case that the work done would be given by the above formula. Thus from the geometric description of the dot product given above, the work equals

$$|\mathbf{F}||\mathbf{p}_2-\mathbf{p}_1|\cos\theta = \mathbf{F}\cdot(\mathbf{p}_2-\mathbf{p}_1).$$

This explains the following definition.

Definition 5.2.7 *Let $\mathbf{F}$ be a force acting on an object which moves from the point, $\mathbf{p}_1$ to the point $\mathbf{p}_2$. Then the **work** done on the object by the given force equals $\mathbf{F}\cdot(\mathbf{p}_2-\mathbf{p}_1)$.*

The concept of writing a given vector, $\mathbf{F}$ in terms of two vectors, one which is parallel to a given vector, $\mathbf{D}$ and the other which is perpendicular can also be explained with no reliance on trigonometry, completely in terms of the algebraic properties of the dot product. As before, this is mathematically more significant than any approach involving geometry or trigonometry because it extends to more interesting situations. This is done next.

Theorem 5.2.8 *Let $\mathbf{F}$ and $\mathbf{D}$ be nonzero vectors. Then there exist unique vectors $\mathbf{F}_{||}$ and $\mathbf{F}_{\perp}$ such that*

$$\mathbf{F} = \mathbf{F}_{||} + \mathbf{F}_{\perp} \tag{5.13}$$

where $\mathbf{F}_{||}$ is a scalar multiple of $\mathbf{D}$, also referred to as

$$\mathrm{proj}_{\mathbf{D}}(\mathbf{F}),$$

*and $\mathbf{F}_{\perp}\cdot\mathbf{D}=0$. The vector $\mathrm{proj}_{\mathbf{D}}(\mathbf{F})$ is called the **projection** of $\mathbf{F}$ onto $\mathbf{D}$.*

Proof: Suppose 5.13 and $\mathbf{F}_{||}=\alpha\mathbf{D}$. Taking the dot product of both sides with $\mathbf{D}$ and using $\mathbf{F}_{\perp}\cdot\mathbf{D}=0$, this yields

$$\mathbf{F}\cdot\mathbf{D} = \alpha|\mathbf{D}|^2$$

which requires $\alpha = \mathbf{F}\cdot\mathbf{D}/|\mathbf{D}|^2$. Thus there can be no more than one vector, $\mathbf{F}_{||}$. It follows $\mathbf{F}_{\perp}$ must equal $\mathbf{F}-\mathbf{F}_{||}$. This verifies there can be no more than one choice for both $\mathbf{F}_{||}$ and $\mathbf{F}_{\perp}$.

Now let

$$\mathbf{F}_{||} \equiv \frac{\mathbf{F}\cdot\mathbf{D}}{|\mathbf{D}|^2}\mathbf{D}$$

and let

$$\mathbf{F}_{\perp} = \mathbf{F}-\mathbf{F}_{||} = \mathbf{F}-\frac{\mathbf{F}\cdot\mathbf{D}}{|\mathbf{D}|^2}\mathbf{D}$$

Then $\mathbf{F}_{||}=\alpha\,\mathbf{D}$ where $\alpha=\frac{\mathbf{F}\cdot\mathbf{D}}{|\mathbf{D}|^2}$. It only remains to verify $\mathbf{F}_{\perp}\cdot\mathbf{D}=0$. But

$$\begin{aligned}\mathbf{F}_{\perp}\cdot\mathbf{D} &= \mathbf{F}\cdot\mathbf{D}-\frac{\mathbf{F}\cdot\mathbf{D}}{|\mathbf{D}|^2}\mathbf{D}\cdot\mathbf{D}\\ &= \mathbf{F}\cdot\mathbf{D}-\mathbf{F}\cdot\mathbf{D}=0.\end{aligned}$$

This proves the theorem.

Example 5.2.9 *Let* $\mathbf{F} = 2\mathbf{i}+7\mathbf{j}-3\mathbf{k}$ *Newtons. Find the work done by this force in moving from the point* $(1,2,3)$ *to the point* $(-9,-3,4)$ *along the straight line segment joining these points where distances are measured in meters.*

According to the definition, this work is

$$\begin{aligned}(2\mathbf{i}+7\mathbf{j}-3\mathbf{k})\cdot(-10\mathbf{i}-5\mathbf{j}+\mathbf{k}) &= -20+(-35)+(-3)\\ &= -58 \text{ Newton meters.}\end{aligned}$$

Note that if the force had been given in pounds and the distance had been given in feet, the units on the work would have been foot pounds. In general, work has units equal to units of a force times units of a length. Instead of writing Newton meter, people write joule because a joule is by definition a Newton meter. That word is pronounced "jewel" and it is the unit of work in the metric system of units. Also be sure you observe that the work done by the force can be negative as in the above example. In fact, work can be either positive, negative, or zero. You just have to do the computations to find out.

Example 5.2.10 *Find* $\mathrm{proj}_{\mathbf{u}}(\mathbf{v})$ *if* $\mathbf{u} = 2\mathbf{i}+3\mathbf{j}-4\mathbf{k}$ *and* $\mathbf{v} = \mathbf{i}-2\mathbf{j}+\mathbf{k}$.

From the above discussion in Theorem 5.2.8, this is just

$$\begin{aligned}&\frac{1}{4+9+16}(\mathbf{i}-2\mathbf{j}+\mathbf{k})\cdot(2\mathbf{i}+3\mathbf{j}-4\mathbf{k})(2\mathbf{i}+3\mathbf{j}-4\mathbf{k})\\ = \quad &\frac{-8}{29}(2\mathbf{i}+3\mathbf{j}-4\mathbf{k}) = -\frac{16}{29}\mathbf{i}-\frac{24}{29}\mathbf{j}+\frac{32}{29}\mathbf{k}.\end{aligned}$$

Example 5.2.11 *Suppose* $\mathbf{a}$, *and* $\mathbf{b}$ *are vectors and* $\mathbf{b}_\perp = \mathbf{b}-\mathrm{proj}_{\mathbf{a}}(\mathbf{b})$. *What is the magnitude of* $\mathbf{b}_\perp$ *in terms of the included angle?*

$$\begin{aligned}|\mathbf{b}_\perp|^2 &= (\mathbf{b}-\mathrm{proj}_{\mathbf{a}}(\mathbf{b}))\cdot(\mathbf{b}-\mathrm{proj}_{\mathbf{a}}(\mathbf{b}))\\ &= \left(\mathbf{b}-\frac{\mathbf{b}\cdot\mathbf{a}}{|\mathbf{a}|^2}\mathbf{a}\right)\cdot\left(\mathbf{b}-\frac{\mathbf{b}\cdot\mathbf{a}}{|\mathbf{a}|^2}\mathbf{a}\right)\\ &= |\mathbf{b}|^2-2\frac{(\mathbf{b}\cdot\mathbf{a})^2}{|\mathbf{a}|^2}+\left(\frac{\mathbf{b}\cdot\mathbf{a}}{|\mathbf{a}|^2}\right)^2|\mathbf{a}|^2\\ &= |\mathbf{b}|^2\left(1-\frac{(\mathbf{b}\cdot\mathbf{a})^2}{|\mathbf{a}|^2|\mathbf{b}|^2}\right)\\ &= |\mathbf{b}|^2\left(1-\cos^2\theta\right) = |\mathbf{b}|^2\sin^2(\theta)\end{aligned}$$

where θ is the included angle between $\mathbf{a}$ and $\mathbf{b}$ which is less than π radians. Therefore, taking square roots,

$$|\mathbf{b}_\perp| = |\mathbf{b}|\sin\theta.$$

5.2.3 The Dot Product And Distance In $\mathbb{C}^n$

It is necessary to give a generalization of the dot product for vectors in $\mathbb{C}^n$. This definition reduces to the usual one in the case the components of the vector are real.

Definition 5.2.12 *Let* $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$. *Thus* $\mathbf{x} = (x_1, \cdots, x_n)$ *where each* $x_k \in \mathbb{C}$ *and a similar formula holding for* $\mathbf{y}$. *Then the dot product of these two vectors is defined to be*

$$\mathbf{x} \cdot \mathbf{y} \equiv \sum_j x_j \overline{y_j} \equiv x_1 \overline{y_1} + \cdots + x_n \overline{y_n}.$$

Notice how you put the conjugate on the entries of the vector, $\mathbf{y}$. It makes no difference if the vectors happen to be real vectors but with complex vectors you must do it this way. The reason for this is that when you take the dot product of a vector with itself, you want to get the square of the length of the vector, a positive number. Placing the conjugate on the components of $\mathbf{y}$ in the above definition assures this will take place. Thus

$$\mathbf{x} \cdot \mathbf{x} = \sum_j x_j \overline{x_j} = \sum_j |x_j|^2 \geq 0.$$

If you didn't place a conjugate as in the above definition, things wouldn't work out correctly. For example,

$$(1+i)^2 + 2^2 = 4 + 2i$$

and this is not a positive number.

The following properties of the dot product follow immediately from the definition and you should verify each of them.

Properties of the dot product:

1. $\mathbf{u} \cdot \mathbf{v} = \overline{\mathbf{v} \cdot \mathbf{u}}$.
2. If a, b are numbers and $\mathbf{u}, \mathbf{v}, \mathbf{z}$ are vectors then $(a\mathbf{u} + b\mathbf{v}) \cdot \mathbf{z} = a(\mathbf{u} \cdot \mathbf{z}) + b(\mathbf{v} \cdot \mathbf{z})$.
3. $\mathbf{u} \cdot \mathbf{u} \geq 0$ and it equals 0 if and only if $\mathbf{u} = \mathbf{0}$.

Note this implies $(\mathbf{x} \cdot \alpha\mathbf{y}) = \overline{\alpha}(\mathbf{x} \cdot \mathbf{y})$ because

$$(\mathbf{x} \cdot \alpha\mathbf{y}) = \overline{(\alpha\mathbf{y} \cdot \mathbf{x})} = \overline{\alpha(\mathbf{y} \cdot \mathbf{x})} = \overline{\alpha}(\mathbf{x} \cdot \mathbf{y})$$

The norm is defined in the usual way.

Definition 5.2.13 *For* $\mathbf{x} \in \mathbb{C}^n$,

$$|\mathbf{x}| \equiv \left(\sum_{k=1}^n |x_k|^2 \right)^{1/2} = (\mathbf{x} \cdot \mathbf{x})^{1/2}$$

Here is a fundamental inequality called the **Cauchy Schwarz inequality** which is stated here in $\mathbb{C}^n$. First here is a simple lemma.

Lemma 5.2.14 *If* $z \in \mathbb{C}$ *there exists* $\theta \in \mathbb{C}$ *such that* $\theta z = |z|$ *and* $|\theta| = 1$.

Proof: Let $\theta = 1$ if $z = 0$ and otherwise, let $\theta = \dfrac{\overline{z}}{|z|}$. Recall that for $z = x+iy, \overline{z} = x-iy$ and $\overline{z}z = |z|^2$.

I will give a proof of this important inequality which depends only on the above list of properties of the dot product. It will be slightly different than the earlier proof.

Theorem 5.2.15 *(Cauchy Schwarz)The following inequality holds for* $\mathbf{x}$ *and* $\mathbf{y} \in \mathbb{C}^n$.

$$|(\mathbf{x} \cdot \mathbf{y})| \leq (\mathbf{x} \cdot \mathbf{x})^{1/2} (\mathbf{y} \cdot \mathbf{y})^{1/2} \tag{5.14}$$

Equality holds in this inequality if and only if one vector is a multiple of the other.

Proof: Let $\theta \in \mathbb{C}$ such that $|\theta| = 1$ and

$$\theta (\mathbf{x} \cdot \mathbf{y}) = |(\mathbf{x} \cdot \mathbf{y})|$$

Consider $p(t) \equiv \left(\mathbf{x} + \overline{\theta} t\mathbf{y}, \mathbf{x} + t\overline{\theta}\mathbf{y}\right)$ where $t \in \mathbb{R}$. Then from the above list of properties of the dot product,

$$\begin{aligned} 0 &\leq p(t) = (\mathbf{x} \cdot \mathbf{x}) + t\theta (\mathbf{x} \cdot \mathbf{y}) + t\overline{\theta} (\mathbf{y} \cdot \mathbf{x}) + t^2 (\mathbf{y} \cdot \mathbf{y}) \\ &= (\mathbf{x} \cdot \mathbf{x}) + t\theta (\mathbf{x} \cdot \mathbf{y}) + t\overline{\theta(\mathbf{x} \cdot \mathbf{y})} + t^2 (\mathbf{y} \cdot \mathbf{y}) \\ &= (\mathbf{x} \cdot \mathbf{x}) + 2t \operatorname{Re} (\theta (\mathbf{x} \cdot \mathbf{y})) + t^2 (\mathbf{y} \cdot \mathbf{y}) \\ &= (\mathbf{x} \cdot \mathbf{x}) + 2t |(\mathbf{x} \cdot \mathbf{y})| + t^2 (\mathbf{y} \cdot \mathbf{y}) \end{aligned} \tag{5.15}$$

and this must hold for all $t \in \mathbb{R}$. Therefore, if $(\mathbf{y} \cdot \mathbf{y}) = 0$ it must be the case that $|(\mathbf{x} \cdot \mathbf{y})| = 0$ also since otherwise the above inequality would be violated. Therefore, in this case,

$$|(\mathbf{x} \cdot \mathbf{y})| \leq (\mathbf{x} \cdot \mathbf{x})^{1/2} (\mathbf{y} \cdot \mathbf{y})^{1/2}.$$

On the other hand, if $(\mathbf{y} \cdot \mathbf{y}) \neq 0$, then $p(t) \geq 0$ for all t means the graph of $y = p(t)$ is a parabola which opens up and it either has exactly one real zero in the case its vertex touches the t axis or it has no real zeros. From the quadratic formula this happens exactly when

$$4 |(\mathbf{x} \cdot \mathbf{y})|^2 - 4 (\mathbf{x} \cdot \mathbf{x}) (\mathbf{y} \cdot \mathbf{y}) \leq 0$$

which is equivalent to 5.14.

It is clear from a computation that if one vector is a scalar multiple of the other that equality holds in 5.14. Conversely, suppose equality does hold. Then this is equivalent to saying $4 |(\mathbf{x} \cdot \mathbf{y})|^2 - 4 (\mathbf{x} \cdot \mathbf{x}) (\mathbf{y} \cdot \mathbf{y}) = 0$ and so from the quadratic formula, there exists one real zero to $p(t) = 0$. Call it t_0. Then

$$p(t_0) \equiv \left(\mathbf{x} + \overline{\theta} t_0 \mathbf{y}, \mathbf{x} + t_0 \overline{\theta}\mathbf{y}\right) = \left|\mathbf{x} + \overline{\theta} t\mathbf{y}\right|^2 = 0$$

and so $\mathbf{x} = -\overline{\theta} t_0 \mathbf{y}$. This proves the theorem.

Note that I only used part of the above properties of the dot product. It was not necessary to use the one which says that if $(\mathbf{x} \cdot \mathbf{x}) = 0$ then $\mathbf{x} = \mathbf{0}$.

By analogy to the case of $\mathbb{R}^n$, length or magnitude of vectors in $\mathbb{C}^n$ can be defined.

Definition 5.2.16 *Let $\mathbf{z} \in \mathbb{C}^n$. Then $|\mathbf{z}| \equiv (\mathbf{z} \cdot \mathbf{z})^{1/2}$.*

Theorem 5.2.17 *For length defined in Definition 5.2.16, the following hold.*

$$|\mathbf{z}| \geq 0 \text{ and } |\mathbf{z}| = 0 \text{ if and only if } \mathbf{z} = \mathbf{0} \tag{5.16}$$

$$\text{If } \alpha \text{ is a scalar, } |\alpha\mathbf{z}| = |\alpha| |\mathbf{z}| \tag{5.17}$$

$$|\mathbf{z} + \mathbf{w}| \leq |\mathbf{z}| + |\mathbf{w}|. \tag{5.18}$$

Proof: The first two claims are left as exercises. To establish the third, you use the same argument which was used in $\mathbb{R}^n$.

$$\begin{aligned} |\mathbf{z} + \mathbf{w}|^2 &= (\mathbf{z} + \mathbf{w}, \mathbf{z} + \mathbf{w}) \\ &= \mathbf{z} \cdot \mathbf{z} + \mathbf{w} \cdot \mathbf{w} + \mathbf{w} \cdot \mathbf{z} + \mathbf{z} \cdot \mathbf{w} \\ &= |\mathbf{z}|^2 + |\mathbf{w}|^2 + 2 \operatorname{Re} \mathbf{w} \cdot \mathbf{z} \\ &\leq |\mathbf{z}|^2 + |\mathbf{w}|^2 + 2 |\mathbf{w} \cdot \mathbf{z}| \\ &\leq |\mathbf{z}|^2 + |\mathbf{w}|^2 + 2 |\mathbf{w}| |\mathbf{z}| = (|\mathbf{z}| + |\mathbf{w}|)^2. \end{aligned}$$

5.3 Exercises With Answers

1. Find the angle between the vectors $3\mathbf{i}-\mathbf{j}-\mathbf{k}$ and $\mathbf{i}+4\mathbf{j}+2\mathbf{k}$.

   $\cos\theta = \frac{3-4-2}{\sqrt{9+1+1}\sqrt{1+16+4}} = -.19739$. Therefore, you have to solve the equation $\cos\theta = -.19739$, Solution is : $\theta = 1.7695$ radians. You need to use a calculator or table to solve this.

2. Find $\mathrm{proj}_{\mathbf{u}}(\mathbf{v})$ where $\mathbf{v}=(1,3,-2)$ and $\mathbf{u}=(1,2,3)$.

   Remember to find this you take $\frac{\mathbf{v}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$. Thus the answer is $\frac{1}{14}(1,2,3)$.

3. If $\mathbf{F}$ is a force and $\mathbf{D}$ is a vector, show $\mathrm{proj}_{\mathbf{D}}(\mathbf{F}) = (|\mathbf{F}|\cos\theta)\,\mathbf{u}$ where $\mathbf{u}$ is the unit vector in the direction of $\mathbf{D}$, $\mathbf{u}=\mathbf{D}/|\mathbf{D}|$ and θ is the included angle between the two vectors, $\mathbf{F}$ and $\mathbf{D}$. $|\mathbf{F}|\cos\theta$ is sometimes called the component of the force, $\mathbf{F}$ in the direction, $\mathbf{D}$.

   $\mathrm{proj}_{\mathbf{D}}(\mathbf{F}) = \frac{\mathbf{F}\cdot\mathbf{D}}{\mathbf{D}\cdot\mathbf{D}}\mathbf{D} = |\mathbf{F}||\mathbf{D}|\cos\theta\frac{1}{|\mathbf{D}|^2}\mathbf{D} = |\mathbf{F}|\cos\theta\frac{\mathbf{D}}{|\mathbf{D}|}$.

4. A boy drags a sled for 100 feet along the ground by pulling on a rope which is 40 degrees from the horizontal with a force of 10 pounds. How much work does this force do?

   The component of force is $10\cos\left(\frac{40}{180}\pi\right)$ and it acts for 100 feet so the work done is

   $$10\cos\left(\frac{40}{180}\pi\right)\times 100 = 766.04$$

5. If $\mathbf{a},\mathbf{b}$, and $\mathbf{c}$ are vectors. Show that $(\mathbf{b}+\mathbf{c})_{\perp} = \mathbf{b}_{\perp}+\mathbf{c}_{\perp}$ where $\mathbf{b}_{\perp} = \mathbf{b} - \mathrm{proj}_{\mathbf{a}}(\mathbf{b})$.

6. Find $(1,0,3,4)\cdot(2,7,1,3)$. $(1,0,3,4)\cdot(2,7,1,3) = 17$.

7. Show that $(\mathbf{a}\cdot\mathbf{b}) = \frac{1}{4}\left[|\mathbf{a}+\mathbf{b}|^2 - |\mathbf{a}-\mathbf{b}|^2\right]$.

   This follows from the axioms of the dot product and the definition of the norm. Thus

   $$|\mathbf{a}+\mathbf{b}|^2 = (\mathbf{a}+\mathbf{b},\mathbf{a}+\mathbf{b}) = |\mathbf{a}|^2+|\mathbf{b}|^2+2(\mathbf{a}\cdot\mathbf{b})$$

   Do something similar for $|\mathbf{a}-\mathbf{b}|^2$.

8. Prove from the axioms of the dot product the parallelogram identity, $|\mathbf{a}+\mathbf{b}|^2 + |\mathbf{a}-\mathbf{b}|^2 = 2|\mathbf{a}|^2+2|\mathbf{b}|^2$.

   Use the properties of the dot product and the definition of the norm in terms of the dot product.

9. Let A and be a real $m\times n$ matrix and let $\mathbf{x}\in\mathbb{R}^n$ and $\mathbf{y}\in\mathbb{R}^m$. Show $(A\mathbf{x},\mathbf{y})_{\mathbb{R}^m} = \left(\mathbf{x},A^T\mathbf{y}\right)_{\mathbb{R}^n}$ where $(\cdot,\cdot)_{\mathbb{R}^k}$ denotes the dot product in $\mathbb{R}^k$. In the notation above, $A\mathbf{x}\cdot\mathbf{y} = \mathbf{x}\cdot A^T\mathbf{y}$. Use the definition of matrix multiplication to do this.

   Remember the ij^{th} entry of $A\mathbf{x} = \sum_j A_{ij}x_j$. Therefore,

   $$A\mathbf{x}\cdot\mathbf{y} = \sum_i (A\mathbf{x})_i\, y_i = \sum_i\sum_j A_{ij}x_j y_i.$$

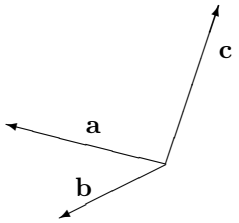
   Recall now that $\left(A^T\right)_{ij} = A_{ji}$. Use this to write a formula for $\left(\mathbf{x},A^T\mathbf{y}\right)_{\mathbb{R}^n}$.

5.4 The Cross Product

The cross product is the other way of multiplying two vectors in $\mathbb{R}^3$. It is very different from the dot product in many ways. First the geometric meaning is discussed and then a description in terms of coordinates is given. Both descriptions of the cross product are important. The geometric description is essential in order to understand the applications to physics and geometry while the coordinate description is the only way to practically compute the cross product.

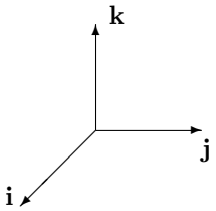
Definition 5.4.1 *Three vectors,* $\mathbf{a}, \mathbf{b}, \mathbf{c}$ *form a right handed system if when you extend the fingers of your right hand along the vector,* $\mathbf{a}$ *and close them in the direction of* $\mathbf{b}$*, the thumb points roughly in the direction of* $\mathbf{c}$*.*

For an example of a right handed system of vectors, see the following picture.



In this picture the vector $\mathbf{c}$ points upwards from the plane determined by the other two vectors. You should consider how a right hand system would differ from a left hand system. Try using your left hand and you will see that the vector, $\mathbf{c}$ would need to point in the opposite direction as it would for a right hand system.

From now on, the vectors, $\mathbf{i}, \mathbf{j}, \mathbf{k}$ will always form a right handed system. To repeat, if you extend the fingers of our right hand along $\mathbf{i}$ and close them in the direction $\mathbf{j}$, the thumb points in the direction of $\mathbf{k}$.

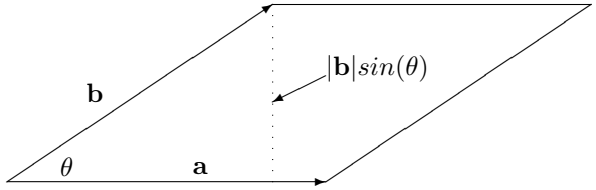


The following is the geometric description of the cross product. It gives both the direction and the magnitude and therefore specifies the vector.

Definition 5.4.2 *Let* $\mathbf{a}$ *and* $\mathbf{b}$ *be two vectors in* $\mathbb{R}^3$*. Then* $\mathbf{a} \times \mathbf{b}$ *is defined by the following two rules.*

1. $|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}|\,|\mathbf{b}| \sin\theta$ *where* θ *is the included angle.*
2. $\mathbf{a} \times \mathbf{b} \cdot \mathbf{a} = 0$, $\mathbf{a} \times \mathbf{b} \cdot \mathbf{b} = 0$, *and* $\mathbf{a}, \mathbf{b}, \mathbf{a} \times \mathbf{b}$ *forms a right hand system.*

Note that $|\mathbf{a}\times\mathbf{b}|$ is the area of the parallelogram determined by $\mathbf{a}$ and $\mathbf{b}$.



The cross product satisfies the following properties.

$$\mathbf{a}\times\mathbf{b} = -(\mathbf{b}\times\mathbf{a}) \ , \ \mathbf{a}\times\mathbf{a} = \mathbf{0}, \tag{5.19}$$

For α a scalar,

$$(\alpha\mathbf{a})\times\mathbf{b} = \alpha(\mathbf{a}\times\mathbf{b}) = \mathbf{a}\times(\alpha\mathbf{b}), \tag{5.20}$$

For $\mathbf{a},\mathbf{b}$, and $\mathbf{c}$ vectors, one obtains the distributive laws,

$$\mathbf{a}\times(\mathbf{b}+\mathbf{c}) = \mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c}, \tag{5.21}$$

$$(\mathbf{b}+\mathbf{c})\times\mathbf{a} = \mathbf{b}\times\mathbf{a}+\mathbf{c}\times\mathbf{a}. \tag{5.22}$$

Formula 5.19 follows immediately from the definition. The vectors $\mathbf{a}\times\mathbf{b}$ and $\mathbf{b}\times\mathbf{a}$ have the same magnitude, $|\mathbf{a}||\mathbf{b}|\sin\theta$, and an application of the right hand rule shows they have opposite direction. Formula 5.20 is also fairly clear. If α is a nonnegative scalar, the direction of $(\alpha\mathbf{a})\times\mathbf{b}$ is the same as the direction of $\mathbf{a}\times\mathbf{b}$,$\alpha(\mathbf{a}\times\mathbf{b})$ and $\mathbf{a}\times(\alpha\mathbf{b})$ while the magnitude is just α times the magnitude of $\mathbf{a}\times\mathbf{b}$ which is the same as the magnitude of $\alpha(\mathbf{a}\times\mathbf{b})$ and $\mathbf{a}\times(\alpha\mathbf{b})$. Using this yields equality in 5.20. In the case where $\alpha<0$, everything works the same way except the vectors are all pointing in the opposite direction and you must multiply by $|\alpha|$ when comparing their magnitudes. The distributive laws are much harder to establish but the second follows from the first quite easily. Thus, assuming the first, and using 5.19,

$$\begin{aligned}(\mathbf{b}+\mathbf{c})\times\mathbf{a} &= -\mathbf{a}\times(\mathbf{b}+\mathbf{c})\\ &= -(\mathbf{a}\times\mathbf{b}+\mathbf{a}\times\mathbf{c})\\ &= \mathbf{b}\times\mathbf{a}+\mathbf{c}\times\mathbf{a}.\end{aligned}$$

A proof of the distributive law is given in a later section for those who are interested.
Now from the definition of the cross product,

$$\begin{array}{ll}\mathbf{i}\times\mathbf{j}=\mathbf{k} & \mathbf{j}\times\mathbf{i}=-\mathbf{k}\\ \mathbf{k}\times\mathbf{i}=\mathbf{j} & \mathbf{i}\times\mathbf{k}=-\mathbf{j}\\ \mathbf{j}\times\mathbf{k}=\mathbf{i} & \mathbf{k}\times\mathbf{j}=-\mathbf{i}\end{array}$$

With this information, the following gives the coordinate description of the cross product.

Proposition 5.4.3 *Let* $\mathbf{a}=a_1\mathbf{i}+a_2\mathbf{j}+a_3\mathbf{k}$ *and* $\mathbf{b}=b_1\mathbf{i}+b_2\mathbf{j}+b_3\mathbf{k}$ *be two vectors. Then*

$$\begin{aligned}\mathbf{a}\times\mathbf{b} = &(a_2b_3-a_3b_2)\,\mathbf{i}+(a_3b_1-a_1b_3)\,\mathbf{j}+\\ &+(a_1b_2-a_2b_1)\,\mathbf{k}.\end{aligned} \tag{5.23}$$

Proof: From the above table and the properties of the cross product listed,

$$\begin{aligned}
&(a_1\mathbf{i}+a_2\mathbf{j}+a_3\mathbf{k})\times(b_1\mathbf{i}+b_2\mathbf{j}+b_3\mathbf{k})=\\
&a_1b_2\mathbf{i}\times\mathbf{j}+a_1b_3\mathbf{i}\times\mathbf{k}+a_2b_1\mathbf{j}\times\mathbf{i}+a_2b_3\mathbf{j}\times\mathbf{k}+\\
&+a_3b_1\mathbf{k}\times\mathbf{i}+a_3b_2\mathbf{k}\times\mathbf{j}\\
&=a_1b_2\mathbf{k}-a_1b_3\mathbf{j}-a_2b_1\mathbf{k}+a_2b_3\mathbf{i}+a_3b_1\mathbf{j}-a_3b_2\mathbf{i}\\
&=(a_2b_3-a_3b_2)\,\mathbf{i}+(a_3b_1-a_1b_3)\,\mathbf{j}+(a_1b_2-a_2b_1)\,\mathbf{k} \qquad (5.24)
\end{aligned}$$

This proves the proposition.

It is probably impossible for most people to remember 5.23. Fortunately, there is a somewhat easier way to remember it. Define the determinant of a 2×2 matrix as follows

$$\left|\begin{array}{cc} a & b \\ c & d \end{array}\right| \equiv ad-bc$$

Then

$$\mathbf{a}\times\mathbf{b}=\left|\begin{array}{ccc} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{array}\right| \qquad (5.25)$$

where you expand the determinant along the top row. This yields

$$\begin{aligned}
&\mathbf{i}\,(-1)^{1+1}\left|\begin{array}{cc} a_2 & a_3 \\ b_2 & b_3 \end{array}\right|+\mathbf{j}\,(-1)^{2+1}\left|\begin{array}{cc} a_1 & a_3 \\ b_1 & b_3 \end{array}\right|+\mathbf{k}\,(-1)^{3+1}\left|\begin{array}{cc} a_1 & a_2 \\ b_1 & b_2 \end{array}\right|\\
&=\mathbf{i}\left|\begin{array}{cc} a_2 & a_3 \\ b_2 & b_3 \end{array}\right|-\mathbf{j}\left|\begin{array}{cc} a_1 & a_3 \\ b_1 & b_3 \end{array}\right|+\mathbf{k}\left|\begin{array}{cc} a_1 & a_2 \\ b_1 & b_2 \end{array}\right|
\end{aligned}$$

Note that to get the scalar which multiplies $\mathbf{i}$ you take the determinant of what is left after deleting the first row and the first column and multiply by $(-1)^{1+1}$ because $\mathbf{i}$ is in the first row and the first column. Then you do the same thing for the $\mathbf{j}$ and $\mathbf{k}$. In the case of the $\mathbf{j}$ there is a minus sign because $\mathbf{j}$ is in the first row and the second column and so$(-1)^{1+2}=-1$ while the $\mathbf{k}$ is multiplied by $(-1)^{3+1}=1$. The above equals

$$(a_2b_3-a_3b_2)\,\mathbf{i}-(a_1b_3-a_3b_1)\,\mathbf{j}+(a_1b_2-a_2b_1)\,\mathbf{k} \qquad (5.26)$$

which is the same as 5.24. There will be much more presented on determinants later. For now, consider this an introduction if you have not seen this topic.

Example 5.4.4 *Find* $(\mathbf{i}-\mathbf{j}+2\mathbf{k})\times(3\mathbf{i}-2\mathbf{j}+\mathbf{k})$.

Use 5.25 to compute this.

$$\begin{aligned}
\left|\begin{array}{ccc} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & -1 & 2 \\ 3 & -2 & 1 \end{array}\right| &= \left|\begin{array}{cc} -1 & 2 \\ -2 & 1 \end{array}\right|\mathbf{i}-\left|\begin{array}{cc} 1 & 2 \\ 3 & 1 \end{array}\right|\mathbf{j}+\left|\begin{array}{cc} 1 & -1 \\ 3 & -2 \end{array}\right|\mathbf{k}\\
&=3\mathbf{i}+5\mathbf{j}+\mathbf{k}.
\end{aligned}$$

Example 5.4.5 *Find the area of the parallelogram determined by the vectors,* $(\mathbf{i}-\mathbf{j}+2\mathbf{k})$ *and* $(3\mathbf{i}-2\mathbf{j}+\mathbf{k})$. *These are the same two vectors in Example 5.4.4.*

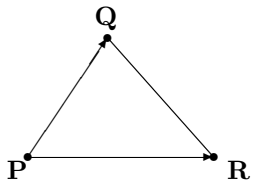
From Example 5.4.4 and the geometric description of the cross product, the area is just the norm of the vector obtained in Example 5.4.4. Thus the area is $\sqrt{9+25+1}=\sqrt{35}$.

Example 5.4.6 *Find the area of the triangle determined by* $(1,2,3)$, $(0,2,5)$, *and* $(5,1,2)$.

This triangle is obtained by connecting the three points with lines. Picking $(1,2,3)$ as a starting point, there are two displacement vectors, $(-1,0,2)$ and $(4,-1,-1)$ such that the given vector added to these displacement vectors gives the other two vectors. The area of the triangle is half the area of the parallelogram determined by $(-1,0,2)$ and $(4,-1,-1)$. Thus $(-1,0,2)\times(4,-1,-1)=(2,7,1)$ and so the area of the triangle is $\frac{1}{2}\sqrt{4+49+1}=\frac{3}{2}\sqrt{6}$.

Observation 5.4.7 *In general, if you have three points (vectors) in* $\mathbb{R}^3, \mathbf{P}, \mathbf{Q}, \mathbf{R}$ *the area of the triangle is given by*

$$\frac{1}{2}\left|(\mathbf{Q}-\mathbf{P})\times(\mathbf{R}-\mathbf{P})\right|.$$

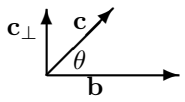


5.4.1 The Distributive Law For The Cross Product

This section gives a proof for 5.21, a fairly difficult topic. It is included here for the interested student. If you are satisfied with taking the distributive law on faith, it is not necessary to read this section. The proof given here is quite clever and follows the one given in [3]. Another approach, based on volumes of parallelepipeds is found in [12] and is discussed a little later.

Lemma 5.4.8 *Let* $\mathbf{b}$ *and* $\mathbf{c}$ *be two vectors. Then* $\mathbf{b}\times\mathbf{c}=\mathbf{b}\times\mathbf{c}_\perp$ *where* $\mathbf{c}_{||}+\mathbf{c}_\perp=\mathbf{c}$ *and* $\mathbf{c}_\perp\cdot\mathbf{b}=0$.

Proof: Consider the following picture.



Now $\mathbf{c}_\perp=\mathbf{c}-\mathbf{c}\cdot\frac{\mathbf{b}}{|\mathbf{b}|}\frac{\mathbf{b}}{|\mathbf{b}|}$ and so $\mathbf{c}_\perp$ is in the plane determined by $\mathbf{c}$ and $\mathbf{b}$. Therefore, from the geometric definition of the cross product, $\mathbf{b}\times\mathbf{c}$ and $\mathbf{b}\times\mathbf{c}_\perp$ have the same direction. Now, referring to the picture,

$$\begin{aligned}|\mathbf{b}\times\mathbf{c}_\perp| &= |\mathbf{b}|\,|\mathbf{c}_\perp| \\ &= |\mathbf{b}|\,|\mathbf{c}|\sin\theta \\ &= |\mathbf{b}\times\mathbf{c}|.\end{aligned}$$

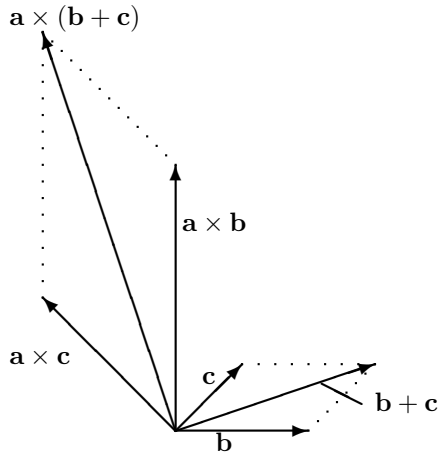
Therefore, $\mathbf{b}\times\mathbf{c}$ and $\mathbf{b}\times\mathbf{c}_\perp$ also have the same magnitude and so they are the same vector.

With this, the proof of the distributive law is in the following theorem.

Theorem 5.4.9 *Let* $\mathbf{a}, \mathbf{b}$, *and* $\mathbf{c}$ *be vectors in* $\mathbb{R}^3$. *Then*

$$\mathbf{a}\times(\mathbf{b}+\mathbf{c}) = \mathbf{a}\times\mathbf{b} + \mathbf{a}\times\mathbf{c} \tag{5.27}$$

Proof: Suppose first that $\mathbf{a}\cdot\mathbf{b} = \mathbf{a}\cdot\mathbf{c} = 0$. Now imagine $\mathbf{a}$ is a vector coming out of the page and let $\mathbf{b}, \mathbf{c}$ and $\mathbf{b}+\mathbf{c}$ be as shown in the following picture.



Then $\mathbf{a}\times\mathbf{b}, \mathbf{a}\times(\mathbf{b}+\mathbf{c})$, and $\mathbf{a}\times\mathbf{c}$ are each vectors in the same plane, perpendicular to $\mathbf{a}$ as shown. Thus $\mathbf{a}\times\mathbf{c}\cdot\mathbf{c} = 0, \mathbf{a}\times(\mathbf{b}+\mathbf{c})\cdot(\mathbf{b}+\mathbf{c}) = 0$, and $\mathbf{a}\times\mathbf{b}\cdot\mathbf{b} = 0$. This implies that to get $\mathbf{a}\times\mathbf{b}$ you move counterclockwise through an angle of $\pi/2$ radians from the vector, $\mathbf{b}$. Similar relationships exist between the vectors $\mathbf{a}\times(\mathbf{b}+\mathbf{c})$ and $\mathbf{b}+\mathbf{c}$ and the vectors $\mathbf{a}\times\mathbf{c}$ and $\mathbf{c}$. Thus the angle between $\mathbf{a}\times\mathbf{b}$ and $\mathbf{a}\times(\mathbf{b}+\mathbf{c})$ is the same as the angle between $\mathbf{b}+\mathbf{c}$ and $\mathbf{b}$ and the angle between $\mathbf{a}\times\mathbf{c}$ and $\mathbf{a}\times(\mathbf{b}+\mathbf{c})$ is the same as the angle between $\mathbf{c}$ and $\mathbf{b}+\mathbf{c}$. In addition to this, since $\mathbf{a}$ is perpendicular to these vectors,

$$|\mathbf{a}\times\mathbf{b}| = |\mathbf{a}||\mathbf{b}|, |\mathbf{a}\times(\mathbf{b}+\mathbf{c})| = |\mathbf{a}||\mathbf{b}+\mathbf{c}|, \text{ and}$$

$$|\mathbf{a}\times\mathbf{c}| = |\mathbf{a}||\mathbf{c}|.$$

Therefore,

$$\frac{|\mathbf{a}\times(\mathbf{b}+\mathbf{c})|}{|\mathbf{b}+\mathbf{c}|} = \frac{|\mathbf{a}\times\mathbf{c}|}{|\mathbf{c}|} = \frac{|\mathbf{a}\times\mathbf{b}|}{|\mathbf{b}|} = |\mathbf{a}|$$

and so

$$\frac{|\mathbf{a}\times(\mathbf{b}+\mathbf{c})|}{|\mathbf{a}\times\mathbf{c}|} = \frac{|\mathbf{b}+\mathbf{c}|}{|\mathbf{c}|}, \frac{|\mathbf{a}\times(\mathbf{b}+\mathbf{c})|}{|\mathbf{a}\times\mathbf{b}|} = \frac{|\mathbf{b}+\mathbf{c}|}{|\mathbf{b}|}$$

showing the triangles making up the parallelogram on the right and the four sided figure on the left in the above picture are similar. It follows the four sided figure on the left is in fact a parallelogram and this implies the diagonal is the vector sum of the vectors on the sides, yielding 5.27.

Now suppose it is not necessarily the case that $\mathbf{a}\cdot\mathbf{b} = \mathbf{a}\cdot\mathbf{c} = 0$. Then write $\mathbf{b} = \mathbf{b}_{||} + \mathbf{b}_\perp$ where $\mathbf{b}_\perp\cdot\mathbf{a} = 0$. Similarly $\mathbf{c} = \mathbf{c}_{||} + \mathbf{c}_\perp$. By the above lemma and what was just shown,

$$\begin{aligned}\mathbf{a}\times(\mathbf{b}+\mathbf{c}) &= \mathbf{a}\times(\mathbf{b}+\mathbf{c})_\perp \\ &= \mathbf{a}\times(\mathbf{b}_\perp + \mathbf{c}_\perp) \\ &= \mathbf{a}\times\mathbf{b}_\perp + \mathbf{a}\times\mathbf{c}_\perp \\ &= \mathbf{a}\times\mathbf{b} + \mathbf{a}\times\mathbf{c}.\end{aligned}$$

This proves the theorem.

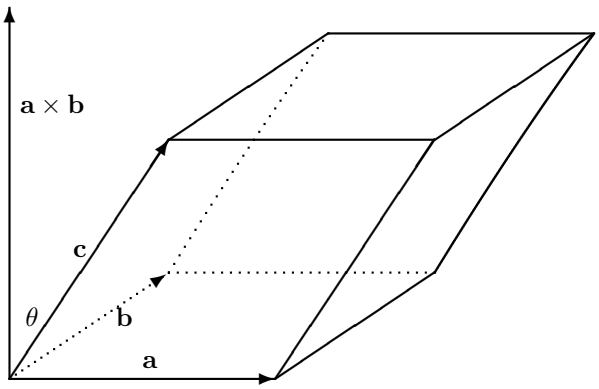
5.4.2 The Box Product

Definition 5.4.10 *A parallelepiped determined by the three vectors,* $\mathbf{a}, \mathbf{b}$, *and* $\mathbf{c}$ *consists of*

$$\{r\mathbf{a}+s\mathbf{b}+t\mathbf{c} : r, s, t \in [0,1]\}.$$

That is, if you pick three numbers, r, s, *and* t *each in* $[0,1]$ *and form* $r\mathbf{a}+s\mathbf{b}+t\mathbf{c}$, *then the collection of all such points is what is meant by the parallelepiped determined by these three vectors.*

The following is a picture of such a thing.



You notice the area of the base of the parallelepiped, the parallelogram determined by the vectors, $\mathbf{a}$ and $\mathbf{b}$ has area equal to $|\mathbf{a}\times\mathbf{b}|$ while the altitude of the parallelepiped is $|\mathbf{c}|\cos\theta$ where θ is the angle shown in the picture between $\mathbf{c}$ and $\mathbf{a}\times\mathbf{b}$. Therefore, the volume of this parallelepiped is the area of the base times the altitude which is just

$$|\mathbf{a}\times\mathbf{b}|\,|\mathbf{c}|\cos\theta = \mathbf{a}\times\mathbf{b}\cdot\mathbf{c}.$$

This expression is known as the box product and is sometimes written as $[\mathbf{a},\mathbf{b},\mathbf{c}]$. You should consider what happens if you interchange the $\mathbf{b}$ with the $\mathbf{c}$ or the $\mathbf{a}$ with the $\mathbf{c}$. You can see geometrically from drawing pictures that this merely introduces a minus sign. In any case the box product of three vectors always equals either the volume of the parallelepiped determined by the three vectors or else minus this volume.

Example 5.4.11 *Find the volume of the parallelepiped determined by the vectors,* $\mathbf{i}+2\mathbf{j}-5\mathbf{k}, \mathbf{i}+3\mathbf{j}-6\mathbf{k}$,$3\mathbf{i}+2\mathbf{j}+3\mathbf{k}$.

According to the above discussion, pick any two of these, take the cross product and then take the dot product of this with the third of these vectors. The result will be either the desired volume or minus the desired volume.

$$\begin{aligned}(\mathbf{i}+2\mathbf{j}-5\mathbf{k})\times(\mathbf{i}+3\mathbf{j}-6\mathbf{k}) &= \begin{vmatrix}\mathbf{i} & \mathbf{j} & \mathbf{k}\\ 1 & 2 & -5\\ 1 & 3 & -6\end{vmatrix}\\ &= 3\mathbf{i}+\mathbf{j}+\mathbf{k}\end{aligned}$$

Now take the dot product of this vector with the third which yields

$$(3\mathbf{i}+\mathbf{j}+\mathbf{k})\cdot(3\mathbf{i}+2\mathbf{j}+3\mathbf{k}) = 9+2+3 = 14.$$

This shows the volume of this parallelepiped is 14 cubic units.

There is a fundamental observation which comes directly from the geometric definitions of the cross product and the dot product.

Lemma 5.4.12 *Let* $\mathbf{a}, \mathbf{b},$ *and* $\mathbf{c}$ *be vectors. Then* $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$.

Proof: This follows from observing that either $(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$ and $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ both give the volume of the parallellepiped or they both give -1 times the volume.

5.4.3 A Proof Of The Distributive Law

Here is another proof of the distributive law for the cross product. Let $\mathbf{x}$ be a vector. From the above observation,

$$\begin{aligned} \mathbf{x} \cdot \mathbf{a} \times (\mathbf{b} + \mathbf{c}) &= (\mathbf{x} \times \mathbf{a}) \cdot (\mathbf{b} + \mathbf{c}) \\ &= (\mathbf{x} \times \mathbf{a}) \cdot \mathbf{b} + (\mathbf{x} \times \mathbf{a}) \cdot \mathbf{c} \\ &= \mathbf{x} \cdot \mathbf{a} \times \mathbf{b} + \mathbf{x} \cdot \mathbf{a} \times \mathbf{c} \\ &= \mathbf{x} \cdot (\mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}) . \end{aligned}$$

Therefore,

$$\mathbf{x} \cdot [\mathbf{a} \times (\mathbf{b} + \mathbf{c}) - (\mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c})] = 0$$

for all $\mathbf{x}$. In particular, this holds for $\mathbf{x} = \mathbf{a} \times (\mathbf{b} + \mathbf{c}) - (\mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c})$ showing that $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$ and this proves the distributive law for the cross product another way.

Observation 5.4.13 *Suppose you have three vectors,* $\mathbf{u} = (a, b, c)$, $\mathbf{v} = (d, e, f)$, *and* $\mathbf{w} = (g, h, i)$. *Then* $\mathbf{u} \cdot \mathbf{v} \times \mathbf{w}$ *is given by the following.*

$$\begin{aligned} \mathbf{u} \cdot \mathbf{v} \times \mathbf{w} &= (a, b, c) \cdot \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ d & e & f \\ g & h & i \end{vmatrix} \\ &= a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} \\ &\equiv \det \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} . \end{aligned}$$

The message is that to take the box product, you can simply take the determinant of the matrix which results by letting the rows be the rectangular components of the given vectors in the order in which they occur in the box product. More will be presented on determinants in the next chapter.

Determinants

6.0.4 Outcomes

A. Evaluate the determinant of a square matrix using by applying
   (a) the cofactor formula or
   (b) row operations.

B. Recall the effects that row operations have on determinants.

C. Recall

1. and verify the following:

   (a) The determinant of a product of matrices is the product of the determinants.
   (b) The determinant of a matrix is equal to the determinant of its transpose.

D. Apply Cramer's Rule to solve a 2×2 or a 3×3 linear system.

E. Use determinants to determine whether a matrix has an inverse.

F. Evaluate the inverse of a matrix using cofactors.

6.1 Basic Techniques And Properties

6.1.1 Cofactors And 2×2 Determinants

Let A be an $n \times n$ matrix. The **determinant** of A, denoted as $\det(A)$ is a number. If the matrix is a 2×2 matrix, this number is very easy to find.

Definition 6.1.1 *Let* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. *Then*

$$\det(A) \equiv ad - cb.$$

The determinant is also often denoted by enclosing the matrix with two vertical lines. Thus

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$

Example 6.1.2 *Find* $\det \begin{pmatrix} 2 & 4 \\ -1 & 6 \end{pmatrix}$.

From the definition this is just $(2)(6) - (-1)(4) = 16$.

Having defined what is meant by the determinant of a 2×2 matrix, what about a 3×3 matrix?

Definition 6.1.3 *Suppose A is a 3×3 matrix. The ij^{th} **minor,** denoted as $\text{minor}(A)_{ij}$, is the determinant of the 2×2 matrix which results from deleting the i^{th} row and the j^{th} column.*

Example 6.1.4 *Consider the matrix,*

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 3 & 2 \\ 3 & 2 & 1 \end{pmatrix}.$$

The $(1,2)$ minor is the determinant of the 2×2 matrix which results when you delete the first row and the second column. This minor is therefore

$$\det \begin{pmatrix} 4 & 2 \\ 3 & 1 \end{pmatrix} = -2.$$

The $(2,3)$ minor is the determinant of the 2×2 matrix which results when you delete the second row and the third column. This minor is therefore

$$\det \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} = -4.$$

Definition 6.1.5 *Suppose A is a 3×3 matrix. The ij^{th} **cofactor** is defined to be $(-1)^{i+j} \times \left(ij^{th} \text{ minor}\right)$. In words, you multiply $(-1)^{i+j}$ times the ij^{th} minor to get the ij^{th} cofactor. The cofactors of a matrix are so important that special notation is appropriate when referring to them. The ij^{th} cofactor of a matrix, A will be denoted by $\text{cof}(A)_{ij}$. It is also convenient to refer to the cofactor of an entry of a matrix as follows. For a_{ij} an entry of the matrix, its cofactor is just $\text{cof}(A)_{ij}$. Thus the cofactor of the ij^{th} entry is just the ij^{th} cofactor.*

Example 6.1.6 *Consider the matrix,*

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 3 & 2 \\ 3 & 2 & 1 \end{pmatrix}.$$

The $(1,2)$ minor is the determinant of the 2×2 matrix which results when you delete the first row and the second column. This minor is therefore

$$\det \begin{pmatrix} 4 & 2 \\ 3 & 1 \end{pmatrix} = -2.$$

It follows

$$\text{cof}(A)_{12} = (-1)^{1+2} \det \begin{pmatrix} 4 & 2 \\ 3 & 1 \end{pmatrix} = (-1)^{1+2}(-2) = 2$$

The $(2,3)$ minor is the determinant of the 2×2 matrix which results when you delete the second row and the third column. This minor is therefore

$$\det \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} = -4.$$

Therefore,

$$\mathrm{cof}\,(A)_{23} = (-1)^{2+3} \det \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix} = (-1)^{2+3}(-4) = 4.$$

Similarly,

$$\mathrm{cof}\,(A)_{22} = (-1)^{2+2} \det \begin{pmatrix} 1 & 3 \\ 3 & 1 \end{pmatrix} = -8.$$

Definition 6.1.7 *The determinant of a 3×3 matrix, A, is obtained by picking a row (column) and taking the product of each entry in that row (column) with its cofactor and adding these up. This process when applied to the i^{th} row (column) is known as expanding the determinant along the i^{th} row (column).*

Example 6.1.8 *Find the determinant of*

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 3 & 2 \\ 3 & 2 & 1 \end{pmatrix}.$$

Here is how it is done by "**expanding along the first column**".

$$\overbrace{1(-1)^{1+1}\begin{vmatrix} 3 & 2 \\ 2 & 1 \end{vmatrix}}^{\mathrm{cof}(A)_{11}} + \overbrace{4(-1)^{2+1}\begin{vmatrix} 2 & 3 \\ 2 & 1 \end{vmatrix}}^{\mathrm{cof}(A)_{21}} + \overbrace{3(-1)^{3+1}\begin{vmatrix} 2 & 3 \\ 3 & 2 \end{vmatrix}}^{\mathrm{cof}(A)_{31}} = 0.$$

You see, we just followed the rule in the above definition. We took the 1 in the first column and multiplied it by its cofactor, the 4 in the first column and multiplied it by its cofactor, and the 3 in the first column and multiplied it by its cofactor. Then we added these numbers together.

You could also expand the determinant along the second row as follows.

$$\overbrace{4(-1)^{2+1}\begin{vmatrix} 2 & 3 \\ 2 & 1 \end{vmatrix}}^{\mathrm{cof}(A)_{21}} + \overbrace{3(-1)^{2+2}\begin{vmatrix} 1 & 3 \\ 3 & 1 \end{vmatrix}}^{\mathrm{cof}(A)_{22}} + \overbrace{2(-1)^{2+3}\begin{vmatrix} 1 & 2 \\ 3 & 2 \end{vmatrix}}^{\mathrm{cof}(A)_{23}} = 0.$$

Observe this gives the same number. You should try expanding along other rows and columns. If you don't make any mistakes, you will always get the same answer.

What about a 4×4 matrix? You know now how to find the determinant of a 3×3 matrix. The pattern is the same.

Definition 6.1.9 *Suppose A is a 4×4 matrix. The ij^{th} **minor** is the determinant of the 3×3 matrix you obtain when you delete the i^{th} row and the j^{th} column. The ij^{th} **cofactor,** $\mathrm{cof}\,(A)_{ij}$ is defined to be $(-1)^{i+j} \times \left(ij^{th} \text{ minor}\right)$. In words, you multiply $(-1)^{i+j}$ times the ij^{th} minor to get the ij^{th} cofactor.*

Definition 6.1.10 *The determinant of a 4×4 matrix, A, is obtained by picking a row (column) and taking the product of each entry in that row (column) with its cofactor and adding these up. This process when applied to the i^{th} row (column) is known as expanding the determinant along the i^{th} row (column).*

Example 6.1.11 *Find* $\det(A)$ *where*

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 4 & 2 & 3 \\ 1 & 3 & 4 & 5 \\ 3 & 4 & 3 & 2 \end{pmatrix}$$

As in the case of a 3×3 matrix, you can expand this along any row or column. Lets pick the third column. $\det(A) =$

$$3(-1)^{1+3}\begin{vmatrix} 5 & 4 & 3 \\ 1 & 3 & 5 \\ 3 & 4 & 2 \end{vmatrix} + 2(-1)^{2+3}\begin{vmatrix} 1 & 2 & 4 \\ 1 & 3 & 5 \\ 3 & 4 & 2 \end{vmatrix} +$$

$$4(-1)^{3+3}\begin{vmatrix} 1 & 2 & 4 \\ 5 & 4 & 3 \\ 3 & 4 & 2 \end{vmatrix} + 3(-1)^{4+3}\begin{vmatrix} 1 & 2 & 4 \\ 5 & 4 & 3 \\ 1 & 3 & 5 \end{vmatrix}.$$

Now you know how to expand each of these 3×3 matrices along a row or a column. If you do so, you will get -12 assuming you make no mistakes. You could expand this matrix along any row or any column and assuming you make no mistakes, you will always get the same thing which is defined to be the determinant of the matrix, A. This method of evaluating a determinant by expanding along a row or a column is called the **method of Laplace expansion**.

Note that each of the four terms above involves three terms consisting of determinants of 2×2 matrices and each of these will need 2 terms. Therefore, there will be $4 \times 3 \times 2 = 24$ terms to evaluate in order to find the determinant using the method of Laplace expansion. Suppose now you have a 10×10 matrix and you follow the above pattern for evaluating determinants. By analogy to the above, there will be $10! = 3,628,800$ terms involved in the evaluation of such a determinant by Laplace expansion along a row or column. This is a lot of terms.

In addition to the difficulties just discussed, you should regard the above claim that you always get the same answer by picking any row or column with considerable skepticism. It is incredible and not at all obvious. However, it requires a little effort to establish it. This is done in the section on the theory of the determinant.

Definition 6.1.12 *Let $A = (a_{ij})$ be an $n \times n$ matrix and suppose the determinant of a $(n-1) \times (n-1)$ matrix has been defined. Then a new matrix called the* ***cofactor matrix****, $\operatorname{cof}(A)$ is defined by $\operatorname{cof}(A) = (c_{ij})$ where to obtain c_{ij} delete the i^{th} row and the j^{th} column of A, take the determinant of the $(n-1) \times (n-1)$ matrix which results, (This is called the ij^{th}* ***minor*** *of A.) and then multiply this number by $(-1)^{i+j}$. Thus $(-1)^{i+j} \times$ (the ij^{th} minor) equals the ij^{th} cofactor. To make the formulas easier to remember, $\operatorname{cof}(A)_{ij}$ will denote the ij^{th} entry of the cofactor matrix.*

With this definition of the cofactor matrix, here is how to define the determinant of an $n \times n$ matrix.

Definition 6.1.13 *Let A be an $n \times n$ matrix where $n \geq 2$ and suppose the determinant of an $(n-1) \times (n-1)$ has been defined. Then*

$$\det(A) = \sum_{j=1}^{n} a_{ij} \operatorname{cof}(A)_{ij} = \sum_{i=1}^{n} a_{ij} \operatorname{cof}(A)_{ij}. \tag{6.1}$$

The first formula consists of expanding the determinant along the i^{th} row and the second expands the determinant along the j^{th} column.

Theorem 6.1.14 *Expanding the $n \times n$ matrix along any row or column always gives the same answer so the above definition is a good definition.*

6.1.2 The Determinant Of A Triangular Matrix

Notwithstanding the difficulties involved in using the method of Laplace expansion, certain types of matrices are very easy to deal with.

Definition 6.1.15 *A matrix M, is upper triangular if $M_{ij} = 0$ whenever $i > j$. Thus such a matrix equals zero below the main diagonal, the entries of the form M_{ii}, as shown.*

$$\begin{pmatrix} * & * & \cdots & * \\ 0 & * & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & * \end{pmatrix}$$

A lower triangular matrix is defined similarly as a matrix for which all entries above the main diagonal are equal to zero.

You should verify the following using the above theorem on Laplace expansion.

Corollary 6.1.16 *Let M be an upper (lower) triangular matrix. Then $\det(M)$ is obtained by taking the product of the entries on the main diagonal.*

Example 6.1.17 *Let*

$$A = \begin{pmatrix} 1 & 2 & 3 & 77 \\ 0 & 2 & 6 & 7 \\ 0 & 0 & 3 & 33.7 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

Find $\det(A)$.

From the above corollary, it suffices to take the product of the diagonal elements. Thus $\det(A) = 1 \times 2 \times 3 \times (-1) = -6$. Without using the corollary, you could expand along the first column. This gives

$$1\begin{vmatrix} 2 & 6 & 7 \\ 0 & 3 & 33.7 \\ 0 & 0 & -1 \end{vmatrix} + 0(-1)^{2+1}\begin{vmatrix} 2 & 3 & 77 \\ 0 & 3 & 33.7 \\ 0 & 0 & -1 \end{vmatrix} +$$
$$0(-1)^{3+1}\begin{vmatrix} 2 & 3 & 77 \\ 2 & 6 & 7 \\ 0 & 0 & -1 \end{vmatrix} + 0(-1)^{4+1}\begin{vmatrix} 2 & 3 & 77 \\ 2 & 6 & 7 \\ 0 & 3 & 33.7 \end{vmatrix}$$

and the only nonzero term in the expansion is

$$1\begin{vmatrix} 2 & 6 & 7 \\ 0 & 3 & 33.7 \\ 0 & 0 & -1 \end{vmatrix}.$$

Now expand this along the first column to obtain

$$1 \times \left(2 \times \begin{vmatrix} 3 & 33.7 \\ 0 & -1 \end{vmatrix} + 0(-1)^{2+1}\begin{vmatrix} 6 & 7 \\ 0 & -1 \end{vmatrix} + 0(-1)^{3+1}\begin{vmatrix} 6 & 7 \\ 3 & 33.7 \end{vmatrix}\right)$$
$$= 1 \times 2 \times \begin{vmatrix} 3 & 33.7 \\ 0 & -1 \end{vmatrix}$$

Next expand this last determinant along the first column to obtain the above equals

$$1 \times 2 \times 3 \times (-1) = -6$$

which is just the product of the entries down the main diagonal of the original matrix.

6.1.3 Properties Of Determinants

There are many properties satisfied by determinants. Some of these properties have to do with row operations. Recall the row operations.

Definition 6.1.18 *The row operations consist of the following*

1. *Switch two rows.*
2. *Multiply a row by a nonzero number.*
3. *Replace a row by a multiple of another row added to itself.*

Theorem 6.1.19 *Let A be an $n \times n$ matrix and let A_1 be a matrix which results from multiplying some row of A by a scalar, c. Then $c \det(A) = \det(A_1)$.*

Example 6.1.20 *Let* $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, A_1 = \begin{pmatrix} 2 & 4 \\ 3 & 4 \end{pmatrix}$. $\det(A) = -2$, $\det(A_1) = -4$.

Theorem 6.1.21 *Let A be an $n \times n$ matrix and let A_1 be a matrix which results from switching two rows of A. Then $\det(A) = -\det(A_1)$. Also, if one row of A is a multiple of another row of A, then $\det(A) = 0$.*

Example 6.1.22 *Let* $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ *and let* $A_1 = \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix}$. $\det A = -2$, $\det(A_1) = 2$.

Theorem 6.1.23 *Let A be an $n \times n$ matrix and let A_1 be a matrix which results from applying row operation 3. That is you replace some row by a multiple of another row added to itself. Then $\det(A) = \det(A_1)$.*

Example 6.1.24 *Let* $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ *and let* $A_1 = \begin{pmatrix} 1 & 2 \\ 4 & 6 \end{pmatrix}$. *Thus the second row of A_1 is one times the first row added to the second row.* $\det(A) = -2$ *and* $\det(A_1) = -2$.

Theorem 6.1.25 *In Theorems 6.1.19 - 6.1.23 you can replace the word, "row" with the word "column".*

There are two other major properties of determinants which do not involve row operations.

Theorem 6.1.26 *Let A and B be two $n \times n$ matrices. Then*

$$\boxed{\det(AB) = \det(A)\det(B).}$$

Also,

$$\boxed{\det(A) = \det\left(A^T\right).}$$

Example 6.1.27 *Compare* $\det(AB)$ *and* $\det(A)\det(B)$ *for*

$$A = \begin{pmatrix} 1 & 2 \\ -3 & 2 \end{pmatrix}, B = \begin{pmatrix} 3 & 2 \\ 4 & 1 \end{pmatrix}.$$

First

$$AB = \begin{pmatrix} 1 & 2 \\ -3 & 2 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 4 & 1 \end{pmatrix} = \begin{pmatrix} 11 & 4 \\ -1 & -4 \end{pmatrix}$$

and so

$$\det(AB) = \det \begin{pmatrix} 11 & 4 \\ -1 & -4 \end{pmatrix} = -40.$$

Now

$$\det(A) = \det \begin{pmatrix} 1 & 2 \\ -3 & 2 \end{pmatrix} = 8$$

and

$$\det(B) = \det \begin{pmatrix} 3 & 2 \\ 4 & 1 \end{pmatrix} = -5.$$

Thus $\det(A)\det(B) = 8 \times (-5) = -40$.

6.1.4 Finding Determinants Using Row Operations

Theorems 6.1.23 - 6.1.25 can be used to find determinants using row operations. As pointed out above, the method of Laplace expansion will not be practical for any matrix of large size. Here is an example in which all the row operations are used.

Example 6.1.28 *Find the determinant of the matrix,*

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 1 & 2 & 3 \\ 4 & 5 & 4 & 3 \\ 2 & 2 & -4 & 5 \end{pmatrix}$$

Replace the second row by (-5) times the first row added to it. Then replace the third row by (-4) times the first row added to it. Finally, replace the fourth row by (-2) times the first row added to it. This yields the matrix,

$$B = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -9 & -13 & -17 \\ 0 & -3 & -8 & -13 \\ 0 & -2 & -10 & -3 \end{pmatrix}$$

and from Theorem 6.1.23, it has the same determinant as A. Now using other row operations, $\det(B) = \left(\frac{-1}{3}\right)\det(C)$ where

$$C = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 0 & 11 & 22 \\ 0 & -3 & -8 & -13 \\ 0 & 6 & 30 & 9 \end{pmatrix}.$$

The second row was replaced by (-3) times the third row added to the second row. By Theorem 6.1.23 this didn't change the value of the determinant. Then the last row was multiplied by (-3). By Theorem 6.1.19 the resulting matrix has a determinant which is (-3) times the determinant of the unmultiplied matrix. Therefore, we multiplied by $-1/3$ to retain the correct value. Now replace the last row with 2 times the third added to it. This does not change the value of the determinant by Theorem 6.1.23. Finally switch

the third and second rows. This causes the determinant to be multiplied by (-1). Thus $\det(C) = -\det(D)$ where

$$D = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -3 & -8 & -13 \\ 0 & 0 & 11 & 22 \\ 0 & 0 & 14 & -17 \end{pmatrix}$$

You could do more row operations or you could note that this can be easily expanded along the first column followed by expanding the 3×3 matrix which results along its first column. Thus

$$\det(D) = 1(-3)\begin{vmatrix} 11 & 22 \\ 14 & -17 \end{vmatrix} = 1485$$

and so $\det(C) = -1485$ and $\det(A) = \det(B) = \left(\frac{-1}{3}\right)(-1485) = 495$.

Example 6.1.29 *Find the determinant of the matrix*

$$\begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & -3 & 2 & 1 \\ 2 & 1 & 2 & 5 \\ 3 & -4 & 1 & 2 \end{pmatrix}$$

Replace the second row by (-1) times the first row added to it. Next take -2 times the first row and add to the third and finally take -3 times the first row and add to the last row. This yields

$$\begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -1 & -1 \\ 0 & -3 & -4 & 1 \\ 0 & -10 & -8 & -4 \end{pmatrix}.$$

By Theorem 6.1.23 this matrix has the same determinant as the original matrix. Remember you can work with the columns also. Take -5 times the last column and add to the second column. This yields

$$\begin{pmatrix} 1 & -8 & 3 & 2 \\ 0 & 0 & -1 & -1 \\ 0 & -8 & -4 & 1 \\ 0 & 10 & -8 & -4 \end{pmatrix}$$

By Theorem 6.1.25 this matrix has the same determinant as the original matrix. Now take (-1) times the third row and add to the top row. This gives.

$$\begin{pmatrix} 1 & 0 & 7 & 1 \\ 0 & 0 & -1 & -1 \\ 0 & -8 & -4 & 1 \\ 0 & 10 & -8 & -4 \end{pmatrix}$$

which by Theorem 6.1.23 has the same determinant as the original matrix. Lets expand it now along the first column. This yields the following for the determinant of the original matrix.

$$\det \begin{pmatrix} 0 & -1 & -1 \\ -8 & -4 & 1 \\ 10 & -8 & -4 \end{pmatrix}$$

which equals

$$8 \det \begin{pmatrix} -1 & -1 \\ -8 & -4 \end{pmatrix} + 10 \det \begin{pmatrix} -1 & -1 \\ -4 & 1 \end{pmatrix} = -82$$

We suggest you do not try to be fancy in using row operations. That is, stick mostly to the one which replaces a row or column with a multiple of another row or column added to it. Also note there is no way to check your answer other than working the problem more than one way. To be sure you have gotten it right you must do this.

6.2 Applications

6.2.1 A Formula For The Inverse

The definition of the determinant in terms of Laplace expansion along a row or column also provides a way to give a formula for the inverse of a matrix. Recall the definition of the inverse of a matrix in Definition 4.1.28 on Page 52. Also recall the definition of the cofactor matrix given in Definition 6.1.12 on Page 80. This cofactor matrix was just the matrix which results from replacing the ij^{th} entry of the matrix with the ij^{th} cofactor.

The following theorem says that to find the inverse, take the transpose of the cofactor matrix and divide by the determinant. The transpose of the cofactor matrix is called the **adjugate** or sometimes the **classical adjoint** of the matrix A. In other words, A^{-1} is equal to one divided by the determinant of A times the adjugate matrix of A. This is what the following theorem says with more precision.

Theorem 6.2.1 *A^{-1} exists if and only if $\det(A) \neq 0$. If $\det(A) \neq 0$, then $A^{-1} = \left(a_{ij}^{-1}\right)$ where*

$$a_{ij}^{-1} = \det(A)^{-1} \operatorname{cof}(A)_{ji}$$

for $\operatorname{cof}(A)_{ij}$ the ij^{th} cofactor of A.

Example 6.2.2 *Find the inverse of the matrix,*

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 0 & 1 \\ 1 & 2 & 1 \end{pmatrix}$$

First find the determinant of this matrix. Using Theorems 6.1.23 - 6.1.25 on Page 82, the determinant of this matrix equals the determinant of the matrix,

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & -6 & -8 \\ 0 & 0 & -2 \end{pmatrix}$$

which equals 12. The cofactor matrix of A is

$$\begin{pmatrix} -2 & -2 & 6 \\ 4 & -2 & 0 \\ 2 & 8 & -6 \end{pmatrix}.$$

Each entry of A was replaced by its cofactor. Therefore, from the above theorem, the inverse of A should equal

$$\frac{1}{12}\begin{pmatrix} -2 & -2 & 6 \\ 4 & -2 & 0 \\ 2 & 8 & -6 \end{pmatrix}^T = \begin{pmatrix} -\frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ -\frac{1}{6} & -\frac{1}{6} & \frac{2}{3} \\ \frac{1}{2} & 0 & -\frac{1}{2} \end{pmatrix}.$$

Does it work? You should check to see if it does. When the matrices are multiplied

$$\begin{pmatrix} -\frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ -\frac{1}{6} & -\frac{1}{6} & \frac{2}{3} \\ \frac{1}{2} & 0 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 3 & 0 & 1 \\ 1 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and so it is correct.

Example 6.2.3 *Find the inverse of the matrix,*

$$A = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ -\frac{1}{6} & \frac{1}{3} & -\frac{1}{2} \\ -\frac{5}{6} & \frac{2}{3} & -\frac{1}{2} \end{pmatrix}$$

First find its determinant. This determinant is $\frac{1}{6}$. The inverse is therefore equal to

$$6\begin{pmatrix} \begin{vmatrix} \frac{1}{3} & -\frac{1}{2} \\ \frac{2}{3} & -\frac{1}{2} \end{vmatrix} & -\begin{vmatrix} -\frac{1}{6} & -\frac{1}{2} \\ -\frac{5}{6} & -\frac{1}{2} \end{vmatrix} & \begin{vmatrix} -\frac{1}{6} & \frac{1}{3} \\ -\frac{5}{6} & \frac{2}{3} \end{vmatrix} \\ -\begin{vmatrix} 0 & \frac{1}{2} \\ \frac{2}{3} & -\frac{1}{2} \end{vmatrix} & \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{5}{6} & -\frac{1}{2} \end{vmatrix} & -\begin{vmatrix} \frac{1}{2} & 0 \\ -\frac{5}{6} & \frac{2}{3} \end{vmatrix} \\ \begin{vmatrix} 0 & \frac{1}{2} \\ \frac{1}{3} & -\frac{1}{2} \end{vmatrix} & -\begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{6} & -\frac{1}{2} \end{vmatrix} & \begin{vmatrix} \frac{1}{2} & 0 \\ -\frac{1}{6} & \frac{1}{3} \end{vmatrix} \end{pmatrix}^T.$$

Expanding all the 2×2 determinants this yields

$$6\begin{pmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{3} & \frac{1}{6} & -\frac{1}{3} \\ -\frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}^T = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 1 & 1 \\ 1 & -2 & 1 \end{pmatrix}$$

Always check your work.

$$\begin{pmatrix} 1 & 2 & -1 \\ 2 & 1 & 1 \\ 1 & -2 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ -\frac{1}{6} & \frac{1}{3} & -\frac{1}{2} \\ -\frac{5}{6} & \frac{2}{3} & -\frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and so we got it right. If the result of multiplying these matrices had been something other than the identity matrix, you would know there was an error. When this happens, you need to search for the mistake if you am interested in getting the right answer. A common mistake is to forget to take the transpose of the cofactor matrix.

Proof of Theorem 6.2.1: From the definition of the determinant in terms of expansion along a column, and letting $(a_{ir}) = A$, if $\det(A) \neq 0$,

$$\sum_{i=1}^{n} a_{ir} \operatorname{cof}(A)_{ir} \det(A)^{-1} = \det(A) \det(A)^{-1} = 1.$$

Now consider

$$\sum_{i=1}^{n} a_{ir} \operatorname{cof}(A)_{ik} \det(A)^{-1}$$

when $k \neq r$. Replace the k^{th} column with the r^{th} column to obtain a matrix, B_k whose determinant equals zero by Theorem 6.1.21. However, expanding this matrix, B_k along the k^{th} column yields

$$0 = \det(B_k) \det(A)^{-1} = \sum_{i=1}^{n} a_{ir} \operatorname{cof}(A)_{ik} \det(A)^{-1}$$

Summarizing,

$$\sum_{i=1}^{n} a_{ir} \operatorname{cof}(A)_{ik} \det(A)^{-1} = \delta_{rk} \equiv \begin{cases} 1 \text{ if } r = k \\ 0 \text{ if } r \neq k \end{cases}.$$

Now

$$\sum_{i=1}^{n} a_{ir} \operatorname{cof}(A)_{ik} = \sum_{i=1}^{n} a_{ir} \operatorname{cof}(A)^{T}_{ki}$$

which is the kr^{th} entry of $\operatorname{cof}(A)^{T} A$. Therefore,

$$\frac{\operatorname{cof}(A)^{T}}{\det(A)} A = I. \tag{6.2}$$

Using the other formula in Definition 6.1.13, and similar reasoning,

$$\sum_{j=1}^{n} a_{rj} \operatorname{cof}(A)_{kj} \det(A)^{-1} = \delta_{rk}$$

Now

$$\sum_{j=1}^{n} a_{rj} \operatorname{cof}(A)_{kj} = \sum_{j=1}^{n} a_{rj} \operatorname{cof}(A)^{T}_{jk}$$

which is the rk^{th} entry of $A \operatorname{cof}(A)^{T}$. Therefore,

$$A \frac{\operatorname{cof}(A)^{T}}{\det(A)} = I, \tag{6.3}$$

and it follows from 6.2 and 6.3 that $A^{-1} = \left(a_{ij}^{-1}\right)$, where

$$a_{ij}^{-1} = \operatorname{cof}(A)_{ji} \det(A)^{-1}.$$

In other words,

$$A^{-1} = \frac{\operatorname{cof}(A)^{T}}{\det(A)}.$$

Now suppose A^{-1} exists. Then by Theorem 6.1.26,

$$1 = \det(I) = \det\left(AA^{-1}\right) = \det(A)\det\left(A^{-1}\right)$$

so $\det(A) \neq 0$. This proves the theorem.

This way of finding inverses is especially useful in the case where it is desired to find the inverse of a matrix whose entries are functions.

Example 6.2.4 *Suppose*

$$A(t) = \begin{pmatrix} e^t & 0 & 0 \\ 0 & \cos t & \sin t \\ 0 & -\sin t & \cos t \end{pmatrix}$$

Show that $A(t)^{-1}$ exists and then find it.

First note $\det(A(t)) = e^t \neq 0$ so $A(t)^{-1}$ exists. The cofactor matrix is

$$C(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^t\cos t & e^t\sin t \\ 0 & -e^t\sin t & e^t\cos t \end{pmatrix}$$

and so the inverse is

$$\frac{1}{e^t}\begin{pmatrix} 1 & 0 & 0 \\ 0 & e^t\cos t & e^t\sin t \\ 0 & -e^t\sin t & e^t\cos t \end{pmatrix}^T = \begin{pmatrix} e^{-t} & 0 & 0 \\ 0 & \cos t & -\sin t \\ 0 & \sin t & \cos t \end{pmatrix}.$$

6.2.2 Cramer's Rule

This formula for the inverse also implies a famous procedure known as **Cramer's rule**. Cramer's rule gives a formula for the solutions, $\mathbf{x}$, to a system of equations, $A\mathbf{x} = \mathbf{y}$ in the special case that A is a square matrix. Note this rule does not apply if you have a system of equations in which there is a different number of equations than variables.

In case you are solving a system of equations, $A\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$, it follows that if A^{-1} exists,

$$\mathbf{x} = \left(A^{-1}A\right)\mathbf{x} = A^{-1}(A\mathbf{x}) = A^{-1}\mathbf{y}$$

thus solving the system. Now in the case that A^{-1} exists, there is a formula for A^{-1} given above. Using this formula,

$$x_i = \sum_{j=1}^{n} a_{ij}^{-1} y_j = \sum_{j=1}^{n} \frac{1}{\det(A)} \operatorname{cof}(A)_{ji}\, y_j.$$

By the formula for the expansion of a determinant along a column,

$$x_i = \frac{1}{\det(A)} \det\begin{pmatrix} * & \cdots & y_1 & \cdots & * \\ \vdots & & \vdots & & \vdots \\ * & \cdots & y_n & \cdots & * \end{pmatrix},$$

where here the i^{th} column of A is replaced with the column vector, $(y_1 \cdots, y_n)^T$, and the determinant of this modified matrix is taken and divided by $\det(A)$. This formula is known as Cramer's rule.

Procedure 6.2.5 *Suppose A is an $n \times n$ matrix and it is desired to solve the system $A\mathbf{x} = \mathbf{y}, \mathbf{y} = (y_1, \cdots, y_n)^T$ for $\mathbf{x} = (x_1, \cdots, x_n)^T$. Then Cramer's rule says*

$$x_i = \frac{\det A_i}{\det A}$$

where A_i is obtained from A by replacing the i^{th} column of A with the column $(y_1, \cdots, y_n)^T$.

Example 6.2.6 *Find x, y if*

$$\begin{pmatrix} 1 & 2 & 1 \\ 3 & 2 & 1 \\ 2 & -3 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

From Cramer's rule,

$$x = \frac{\begin{vmatrix} 1 & 2 & 1 \\ 2 & 2 & 1 \\ 3 & -3 & 2 \end{vmatrix}}{\begin{vmatrix} 1 & 2 & 1 \\ 3 & 2 & 1 \\ 2 & -3 & 2 \end{vmatrix}} = \frac{1}{2}$$

Now to find y,

$$y = \frac{\begin{vmatrix} 1 & 1 & 1 \\ 3 & 2 & 1 \\ 2 & 3 & 2 \end{vmatrix}}{\begin{vmatrix} 1 & 2 & 1 \\ 3 & 2 & 1 \\ 2 & -3 & 2 \end{vmatrix}} = -\frac{1}{7}$$

$$z = \frac{\begin{vmatrix} 1 & 2 & 1 \\ 3 & 2 & 2 \\ 2 & -3 & 3 \end{vmatrix}}{\begin{vmatrix} 1 & 2 & 1 \\ 3 & 2 & 1 \\ 2 & -3 & 2 \end{vmatrix}} = \frac{11}{14}$$

You see the pattern. For large systems Cramer's rule is less than useful if you want to find an answer. This is because to use it you must evaluate determinants. However, you have no practical way to evaluate determinants for large matrices other than row operations and if you are using row operations, you might just as well use them to solve the system to begin with. It will be a lot less trouble. Nevertheless, there are situations in which Cramer's rule is useful.

Example 6.2.7 *Solve for z if*

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & -e^t \sin t & e^t \cos t \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ t \\ t^2 \end{pmatrix}$$

You could do it by row operations but it might be easier in this case to use Cramer's rule because the matrix of coefficients does not consist of numbers but of functions. Thus

$$z = \frac{\begin{vmatrix} 1 & 0 & 1 \\ 0 & e^t \cos t & t \\ 0 & -e^t \sin t & t^2 \end{vmatrix}}{\begin{vmatrix} 1 & 0 & 0 \\ 0 & e^t \cos t & e^t \sin t \\ 0 & -e^t \sin t & e^t \cos t \end{vmatrix}} = t\left(\left(\cos t\right) t + \sin t\right) e^{-t}.$$

You end up doing this sort of thing sometimes in ordinary differential equations in the method of variation of parameters.

6.3 Exercises With Answers

1. Find the following determinant by expanding along the second column.

$$\begin{vmatrix} 1 & 3 & 1 \\ 2 & 1 & 5 \\ 2 & 1 & 1 \end{vmatrix}$$

   This is

$$3\left(-1\right)^{2+1}\begin{vmatrix} 2 & 5 \\ 2 & 1 \end{vmatrix} + 1\left(-1\right)^{1+1}\begin{vmatrix} 1 & 1 \\ 2 & 1 \end{vmatrix} + 1\left(-1\right)^{3+2}\begin{vmatrix} 1 & 1 \\ 2 & 5 \end{vmatrix} = 20.$$

2. Compute the determinant by cofactor expansion. Pick the easiest row or column to use.

$$\begin{vmatrix} 2 & 0 & 0 & 1 \\ 2 & 1 & 1 & 0 \\ 0 & 0 & 0 & 3 \\ 2 & 3 & 3 & 1 \end{vmatrix}$$

   You ought to use the third row. This yields

$$3\begin{vmatrix} 2 & 0 & 0 \\ 2 & 1 & 1 \\ 2 & 3 & 3 \end{vmatrix} = \left(3\right)\left(2\right)\begin{vmatrix} 1 & 1 \\ 3 & 3 \end{vmatrix} = 0.$$

3. Find the determinant using row and column operations.

$$\begin{vmatrix} 5 & 4 & 3 & 2 \\ 3 & 2 & 4 & 3 \\ -1 & 2 & 3 & 3 \\ 2 & 1 & 2 & -2 \end{vmatrix}$$

   Replace the first row by 5 times the third added to it and then replace the second by 3 times the third added to it and then the last by 2 times the third added to it. This yields

$$\begin{vmatrix} 0 & 14 & 18 & 17 \\ 0 & 8 & 13 & 12 \\ -1 & 2 & 3 & 3 \\ 0 & 5 & 8 & 4 \end{vmatrix}$$

Now lets replace the third column by -1 times the last column added to it.

$$\begin{vmatrix} 0 & 14 & 1 & 17 \\ 0 & 8 & 1 & 12 \\ -1 & 2 & 0 & 3 \\ 0 & 5 & 4 & 4 \end{vmatrix}$$

Now replace the top row by -1 times the second added to it and the bottom row by -4 times the second added to it. This yields

$$\begin{vmatrix} 0 & 6 & 0 & 5 \\ 0 & 8 & 1 & 12 \\ -1 & 2 & 0 & 3 \\ 0 & -27 & 0 & -44 \end{vmatrix}. \tag{6.4}$$

This looks pretty good because it has a lot of zeros. Expand along the first column and next along the second,

$$(-1)\begin{vmatrix} 6 & 0 & 5 \\ 8 & 1 & 12 \\ -27 & 0 & -44 \end{vmatrix} = (-1)(1)\begin{vmatrix} 6 & 5 \\ -27 & -44 \end{vmatrix} = 129.$$

Alternatively, you could continue doing row and column operations. Switch the third and first row in 6.4 to obtain

$$-\begin{vmatrix} -1 & 2 & 0 & 3 \\ 0 & 8 & 1 & 12 \\ 0 & 6 & 0 & 5 \\ 0 & -27 & 0 & -44 \end{vmatrix}$$

Next take $9/2$ times the third row and add to the bottom.

$$-\begin{vmatrix} -1 & 2 & 0 & 3 \\ 0 & 8 & 1 & 12 \\ 0 & 6 & 0 & 5 \\ 0 & 0 & 0 & -44+(9/2)5 \end{vmatrix}.$$

Finally, take $-6/8$ times the second row and add to the third.

$$-\begin{vmatrix} -1 & 2 & 0 & 3 \\ 0 & 8 & 1 & 12 \\ 0 & 0 & -6/8 & 5+(-6/8)(12) \\ 0 & 0 & 0 & -44+(9/2)5 \end{vmatrix}.$$

Therefore, since the matrix is now upper triangular, the determinant is

$$-((-1)(8)(-6/8)(-44+(9/2)5)) = 129.$$

4. An operation is done to get from the first matrix to the second. Identify what was done and tell how it will affect the value of the determinant.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

This involved taking the transpose so the determinant of the new matrix is the same as the determinant of the first matrix.

5. Show that for A a 2×2 matrix $\det(aA) = a^2 \det(A)$ where a is a scalar.

   $a^2 \det(A) = a \det(A_1)$ where the first row of A is replaced by a times it to get A_1. Then $a \det(A_1) = A_2$ where A_2 is obtained from A by multiplying both rows by a. In other words, $A_2 = aA$. Thus the conclusion is established.

6. Use Cramer's rule to find y in

$$\begin{array}{c} 2x + 2y + z = 3 \\ 2x - y - z = 2 \\ x + 2z = 1 \end{array}$$

   From Cramer's rule,

$$y = \frac{\begin{vmatrix} 2 & 3 & 1 \\ 2 & 2 & -1 \\ 1 & 1 & 2 \end{vmatrix}}{\begin{vmatrix} 2 & 2 & 1 \\ 2 & -1 & -1 \\ 1 & 0 & 2 \end{vmatrix}} = \frac{5}{13}.$$

7. Here is a matrix,

$$\begin{pmatrix} e^t & e^{-t}\cos t & e^{-t}\sin t \\ e^t & -e^{-t}\cos t - e^{-t}\sin t & -e^{-t}\sin t + e^{-t}\cos t \\ e^t & 2e^{-t}\sin t & -2e^{-t}\cos t \end{pmatrix}$$

   Does there exist a value of t for which this matrix fails to have an inverse? Explain.

$\det \begin{pmatrix} e^t & e^{-t}\cos t & e^{-t}\sin t \\ e^t & -e^{-t}\cos t - e^{-t}\sin t & -e^{-t}\sin t + e^{-t}\cos t \\ e^t & 2e^{-t}\sin t & -2e^{-t}\cos t \end{pmatrix}$
$= 5e^t e^{2(-t)}\cos^2 t + 5e^t e^{2(-t)}\sin^2 t = 5e^{-t}$ which is never equal to zero for any value of t and so there is no value of t for which the matrix has no inverse.

8. Use the formula for the inverse in terms of the cofactor matrix to find if possible the inverse of the matrix

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 6 & 1 \\ 4 & 1 & 1 \end{pmatrix}.$$

   First you need to take the determinant

$$\det \begin{pmatrix} 1 & 2 & 3 \\ 0 & 6 & 1 \\ 4 & 1 & 1 \end{pmatrix} = -59$$

   and so the matrix has an inverse. Now you need to find the cofactor matrix.

$$\begin{pmatrix} \begin{vmatrix} 6 & 1 \\ 1 & 1 \end{vmatrix} & -\begin{vmatrix} 0 & 1 \\ 4 & 1 \end{vmatrix} & \begin{vmatrix} 0 & 6 \\ 4 & 1 \end{vmatrix} \\ -\begin{vmatrix} 2 & 3 \\ 1 & 1 \end{vmatrix} & \begin{vmatrix} 1 & 3 \\ 4 & 1 \end{vmatrix} & -\begin{vmatrix} 1 & 2 \\ 4 & 1 \end{vmatrix} \\ \begin{vmatrix} 2 & 3 \\ 6 & 1 \end{vmatrix} & -\begin{vmatrix} 1 & 3 \\ 0 & 1 \end{vmatrix} & \begin{vmatrix} 1 & 2 \\ 0 & 6 \end{vmatrix} \end{pmatrix}$$

$$= \begin{pmatrix} 5 & 4 & -24 \\ 1 & -11 & 7 \\ -16 & -1 & 6 \end{pmatrix}.$$

Thus the inverse is

$$\frac{1}{-59}\begin{pmatrix} 5 & 4 & -24 \\ 1 & -11 & 7 \\ -16 & -1 & 6 \end{pmatrix}^T$$

$$= \frac{1}{-59}\begin{pmatrix} 5 & 1 & -16 \\ 4 & -11 & -1 \\ -24 & 7 & 6 \end{pmatrix}.$$

If you check this, it does work.

6.4 The Mathematical Theory Of Determinants*

This material is definitely not for the faint of heart. It is only for people who want to see everything proved. It is a fairly complete and unusually elementary treatment of the subject. There will be some repetition between this section and the earlier section on determinants. The main purpose is to give all the missing proofs. Two books which give a good introduction to determinants are Apostol [1] and Rudin [11]. A recent book which also has a good introduction is Baker [2]. Most linear algebra books do not do an honest job presenting this topic.

It is easiest to give a different definition of the determinant which is clearly well defined and then prove the earlier one in terms of Laplace expansion. Let $(i_1, \cdots, i_n)$ be an ordered list of numbers from $\{1, \cdots, n\}$. This means the order is important so $(1,2,3)$ and $(2,1,3)$ are different.

The following Lemma will be essential in the definition of the determinant.

Lemma 6.4.1 *There exists a unique function,* sgn_n *which maps each list of numbers from* $\{1, \cdots, n\}$ *to one of the three numbers,* $0, 1,$ *or* -1 *which also has the following properties.*

$$\operatorname{sgn}_n(1, \cdots, n) = 1 \tag{6.5}$$

$$\operatorname{sgn}_n(i_1, \cdots, p, \cdots, q, \cdots, i_n) = -\operatorname{sgn}_n(i_1, \cdots, q, \cdots, p, \cdots, i_n) \tag{6.6}$$

In words, the second property states that if two of the numbers are switched, the value of the function is multiplied by -1*. Also, in the case where* $n > 1$ *and* $\{i_1, \cdots, i_n\} = \{1, \cdots, n\}$ *so that every number from* $\{1, \cdots, n\}$ *appears in the ordered list,* $(i_1, \cdots, i_n)$*,*

$$\operatorname{sgn}_n(i_1, \cdots, i_{\theta-1}, n, i_{\theta+1}, \cdots, i_n) \equiv$$

$$(-1)^{n-\theta}\operatorname{sgn}_{n-1}(i_1,\cdots,i_{\theta-1},i_{\theta+1},\cdots,i_n) \tag{6.7}$$

where $n = i_\theta$ in the ordered list, $(i_1,\cdots,i_n)$.

Proof: To begin with, it is necessary to show the existence of such a function. This is clearly true if $n = 1$. Define $\operatorname{sgn}_1(1) \equiv 1$ and observe that it works. No switching is possible. In the case where $n = 2$, it is also clearly true. Let $\operatorname{sgn}_2(1,2) = 1$ and $\operatorname{sgn}_2(2,1) = 0$ while $\operatorname{sgn}_2(2,2) = \operatorname{sgn}_2(1,1) = 0$ and verify it works. Assuming such a function exists for n, sgn_{n+1} will be defined in terms of sgn_n. If there are any repeated numbers in $(i_1,\cdots,i_{n+1})$, $\operatorname{sgn}_{n+1}(i_1,\cdots,i_{n+1}) \equiv 0$. If there are no repeats, then $n+1$ appears somewhere in the ordered list. Let θ be the position of the number $n+1$ in the list. Thus, the list is of the form $(i_1,\cdots,i_{\theta-1},n+1,i_{\theta+1},\cdots,i_{n+1})$. From 6.7 it must be that

$$\operatorname{sgn}_{n+1}(i_1,\cdots,i_{\theta-1},n+1,i_{\theta+1},\cdots,i_{n+1}) \equiv$$

$$(-1)^{n+1-\theta}\operatorname{sgn}_n(i_1,\cdots,i_{\theta-1},i_{\theta+1},\cdots,i_{n+1}).$$

It is necessary to verify this satisfies 6.5 and 6.6 with n replaced with $n+1$. The first of these is obviously true because

$$\operatorname{sgn}_{n+1}(1,\cdots,n,n+1) \equiv (-1)^{n+1-(n+1)}\operatorname{sgn}_n(1,\cdots,n) = 1.$$

If there are repeated numbers in $(i_1,\cdots,i_{n+1})$, then it is obvious 6.6 holds because both sides would equal zero from the above definition. It remains to verify 6.6 in the case where there are no numbers repeated in $(i_1,\cdots,i_{n+1})$. Consider

$$\operatorname{sgn}_{n+1}\left(i_1,\cdots,\overset{r}{p},\cdots,\overset{s}{q},\cdots,i_{n+1}\right),$$

where the r above the p indicates the number, p is in the r^{th} position and the s above the q indicates that the number, q is in the s^{th} position. Suppose first that $r < \theta < s$. Then

$$\operatorname{sgn}_{n+1}\left(i_1,\cdots,\overset{r}{p},\cdots,\overset{\theta}{n+1},\cdots,\overset{s}{q},\cdots,i_{n+1}\right) \equiv$$

$$(-1)^{n+1-\theta}\operatorname{sgn}_n\left(i_1,\cdots,\overset{r}{p},\cdots,\overset{s-1}{q},\cdots,i_{n+1}\right)$$

while

$$\operatorname{sgn}_{n+1}\left(i_1,\cdots,\overset{r}{q},\cdots,\overset{\theta}{n+1},\cdots,\overset{s}{p},\cdots,i_{n+1}\right) =$$

$$(-1)^{n+1-\theta}\operatorname{sgn}_n\left(i_1,\cdots,\overset{r}{q},\cdots,\overset{s-1}{p},\cdots,i_{n+1}\right)$$

and so, by induction, a switch of p and q introduces a minus sign in the result. Similarly, if $\theta > s$ or if $\theta < r$ it also follows that 6.6 holds. The interesting case is when $\theta = r$ or $\theta = s$. Consider the case where $\theta = r$ and note the other case is entirely similar.

$$\operatorname{sgn}_{n+1}\left(i_1,\cdots,\overset{r}{n+1},\cdots,\overset{s}{q},\cdots,i_{n+1}\right) =$$

$$(-1)^{n+1-r}\operatorname{sgn}_n\left(i_1,\cdots,\overset{s-1}{q},\cdots,i_{n+1}\right) \tag{6.8}$$

while

$$\operatorname{sgn}_{n+1}\left(i_1,\cdots,\overset{r}{q},\cdots,\overset{s}{n+1},\cdots,i_{n+1}\right) =$$

$$(-1)^{n+1-s}\operatorname{sgn}_n\left(i_1,\cdots,\overset{r}{q},\cdots,i_{n+1}\right). \tag{6.9}$$

By making $s-1-r$ switches, move the q which is in the $s-1^{th}$ position in 6.8 to the r^{th} position in 6.9. By induction, each of these switches introduces a factor of -1 and so

$$\operatorname{sgn}_n\left(i_1,\cdots,\overset{s-1}{q},\cdots,i_{n+1}\right)=(-1)^{s-1-r}\operatorname{sgn}_n\left(i_1,\cdots,\overset{r}{q},\cdots,i_{n+1}\right).$$

Therefore,

$$\begin{aligned}\operatorname{sgn}_{n+1}\left(i_1,\cdots,\overset{r}{n+1},\cdots,\overset{s}{q},\cdots,i_{n+1}\right)&=(-1)^{n+1-r}\operatorname{sgn}_n\left(i_1,\cdots,\overset{s-1}{q},\cdots,i_{n+1}\right)\\&=(-1)^{n+1-r}(-1)^{s-1-r}\operatorname{sgn}_n\left(i_1,\cdots,\overset{r}{q},\cdots,i_{n+1}\right)\\=(-1)^{n+s}\operatorname{sgn}_n\left(i_1,\cdots,\overset{r}{q},\cdots,i_{n+1}\right)&=(-1)^{2s-1}(-1)^{n+1-s}\operatorname{sgn}_n\left(i_1,\cdots,\overset{r}{q},\cdots,i_{n+1}\right)\\&=-\operatorname{sgn}_{n+1}\left(i_1,\cdots,\overset{r}{q},\cdots,\overset{s}{n+1},\cdots,i_{n+1}\right).\end{aligned}$$

This proves the existence of the desired function.

To see this function is unique, note that you can obtain any ordered list of distinct numbers from a sequence of switches. If there exist two functions, f and g both satisfying 6.5 and 6.6, you could start with $f(1,\cdots,n)=g(1,\cdots,n)$ and applying the same sequence of switches, eventually arrive at $f(i_1,\cdots,i_n)=g(i_1,\cdots,i_n)$. If any numbers are repeated, then 6.6 gives both functions are equal to zero for that ordered list. This proves the lemma.

In what follows sgn will often be used rather than sgn_n because the context supplies the appropriate n.

Definition 6.4.2 *Let f be a real valued function which has the set of ordered lists of numbers from $\{1,\cdots,n\}$ as its domain. Define*

$$\sum_{(k_1,\cdots,k_n)} f(k_1\cdots k_n)$$

to be the sum of all the $f(k_1\cdots k_n)$ for all possible choices of ordered lists $(k_1,\cdots,k_n)$ of numbers of $\{1,\cdots,n\}$. For example,

$$\sum_{(k_1,k_2)} f(k_1,k_2)=f(1,2)+f(2,1)+f(1,1)+f(2,2).$$

Definition 6.4.3 *Let $(a_{ij})=A$ denote an $n\times n$ matrix. The determinant of A, denoted by $\det(A)$ is defined by*

$$\det(A)\equiv\sum_{(k_1,\cdots,k_n)}\operatorname{sgn}(k_1,\cdots,k_n)\,a_{1k_1}\cdots a_{nk_n}$$

where the sum is taken over all ordered lists of numbers from $\{1,\cdots,n\}$. Note it suffices to take the sum over only those ordered lists in which there are no repeats because if there are, $\operatorname{sgn}(k_1,\cdots,k_n)=0$ and so that term contributes 0 to the sum.

Let A be an $n\times n$ matrix, $A=(a_{ij})$ and let $(r_1,\cdots,r_n)$ denote an ordered list of n numbers from $\{1,\cdots,n\}$. Let $A(r_1,\cdots,r_n)$ denote the matrix whose k^{th} row is the r_k row of the matrix, A. Thus

$$\det(A(r_1,\cdots,r_n))=\sum_{(k_1,\cdots,k_n)}\operatorname{sgn}(k_1,\cdots,k_n)\,a_{r_1k_1}\cdots a_{r_nk_n}\tag{6.10}$$

and

$$A(1,\cdots,n)=A.$$

Proposition 6.4.4 *Let*

$$(r_1, \cdots, r_n)$$

be an ordered list of numbers from $\{1, \cdots, n\}$. *Then*

$$\operatorname{sgn}(r_1, \cdots, r_n) \det(A)$$

$$= \sum_{(k_1, \cdots, k_n)} \operatorname{sgn}(k_1, \cdots, k_n) a_{r_1 k_1} \cdots a_{r_n k_n} \tag{6.11}$$

$$= \det(A(r_1, \cdots, r_n)). \tag{6.12}$$

Proof: Let $(1, \cdots, n) = (1, \cdots, r, \cdots s, \cdots, n)$ so $r < s$.

$$\det(A(1, \cdots, r, \cdots, s, \cdots, n)) = \tag{6.13}$$

$$\sum_{(k_1, \cdots, k_n)} \operatorname{sgn}(k_1, \cdots, k_r, \cdots, k_s, \cdots, k_n) a_{1k_1} \cdots a_{rk_r} \cdots a_{sk_s} \cdots a_{nk_n},$$

and renaming the variables, calling k_s, k_r and k_r, k_s, this equals

$$= \sum_{(k_1, \cdots, k_n)} \operatorname{sgn}(k_1, \cdots, k_s, \cdots, k_r, \cdots, k_n) a_{1k_1} \cdots a_{rk_s} \cdots a_{sk_r} \cdots a_{nk_n}$$

$$= \sum_{(k_1, \cdots, k_n)} -\operatorname{sgn}\left(k_1, \cdots, \overbrace{k_r, \cdots, k_s}^{\text{These got switched}}, \cdots, k_n\right) a_{1k_1} \cdots a_{sk_r} \cdots a_{rk_s} \cdots a_{nk_n}$$

$$= -\det(A(1, \cdots, s, \cdots, r, \cdots, n)). \tag{6.14}$$

Consequently,

$$\det(A(1, \cdots, s, \cdots, r, \cdots, n)) =$$

$$-\det(A(1, \cdots, r, \cdots, s, \cdots, n)) = -\det(A)$$

Now letting $A(1, \cdots, s, \cdots, r, \cdots, n)$ play the role of A, and continuing in this way, switching pairs of numbers,

$$\det(A(r_1, \cdots, r_n)) = (-1)^p \det(A)$$

where it took p switches to obtain$(r_1, \cdots, r_n)$ from $(1, \cdots, n)$. By Lemma 6.4.1, this implies

$$\det(A(r_1, \cdots, r_n)) = (-1)^p \det(A) = \operatorname{sgn}(r_1, \cdots, r_n) \det(A)$$

and proves the proposition in the case when there are no repeated numbers in the ordered list, $(r_1, \cdots, r_n)$. However, if there is a repeat, say the r^{th} row equals the s^{th} row, then the reasoning of 6.13 -6.14 shows that $A(r_1, \cdots, r_n) = 0$ and also $\operatorname{sgn}(r_1, \cdots, r_n) = 0$ so the formula holds in this case also.

Observation 6.4.5 *There are* $n!$ *ordered lists of distinct numbers from* $\{1, \cdots, n\}$.

To see this, consider n slots placed in order. There are n choices for the first slot. For each of these choices, there are $n-1$ choices for the second. Thus there are $n(n-1)$ ways to fill the first two slots. Then for each of these ways there are $n-2$ choices left for the third slot. Continuing this way, there are $n!$ ordered lists of distinct numbers from $\{1, \cdots, n\}$ as stated in the observation.

With the above, it is possible to give a more symmetric description of the determinant from which it will follow that $\det(A) = \det\left(A^T\right)$.

Corollary 6.4.6 *The following formula for* $\det(A)$ *is valid.*

$$\det(A) = \frac{1}{n!}\cdot$$

$$\sum_{(r_1,\cdots,r_n)} \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(r_1,\cdots,r_n) \operatorname{sgn}(k_1,\cdots,k_n) a_{r_1k_1}\cdots a_{r_nk_n}. \tag{6.15}$$

And also $\det\left(A^T\right) = \det(A)$ *where* A^T *is the transpose of* A*. (Recall that for* $A^T = \left(a_{ij}^T\right)$*,* $a_{ij}^T = a_{ji}$*.)*

Proof: From Proposition 6.4.4, if the r_i are distinct,

$$\det(A) = \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(r_1,\cdots,r_n) \operatorname{sgn}(k_1,\cdots,k_n) a_{r_1k_1}\cdots a_{r_nk_n}.$$

Summing over all ordered lists, $(r_1,\cdots,r_n)$ where the r_i are distinct, (If the r_i are not distinct, $\operatorname{sgn}(r_1,\cdots,r_n) = 0$ and so there is no contribution to the sum.)

$$n!\det(A) =$$

$$\sum_{(r_1,\cdots,r_n)} \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(r_1,\cdots,r_n) \operatorname{sgn}(k_1,\cdots,k_n) a_{r_1k_1}\cdots a_{r_nk_n}.$$

This proves the corollary since the formula gives the same number for A as it does for A^T.

Corollary 6.4.7 *If two rows or two columns in an* $n\times n$ *matrix,* A*, are switched, the determinant of the resulting matrix equals* (-1) *times the determinant of the original matrix. If* A *is an* $n\times n$ *matrix in which two rows are equal or two columns are equal then* $\det(A) = 0$*. Suppose the* i^{th} *row of* A *equals* $(xa_1 + yb_1,\cdots,xa_n + yb_n)$*. Then*

$$\det(A) = x\det(A_1) + y\det(A_2)$$

where the i^{th} *row of* A_1 *is* $(a_1,\cdots,a_n)$ *and the* i^{th} *row of* A_2 *is* $(b_1,\cdots,b_n)$*, all other rows of* A_1 *and* A_2 *coinciding with those of* A*. In other words,* det *is a linear function of each row* A*. The same is true with the word "row" replaced with the word "column".*

Proof: By Proposition 6.4.4 when two rows are switched, the determinant of the resulting matrix is (-1) times the determinant of the original matrix. By Corollary 6.4.6 the same holds for columns because the columns of the matrix equal the rows of the transposed matrix. Thus if A_1 is the matrix obtained from A by switching two columns,

$$\det(A) = \det\left(A^T\right) = -\det\left(A_1^T\right) = -\det(A_1).$$

If A has two equal columns or two equal rows, then switching them results in the same matrix. Therefore, $\det(A) = -\det(A)$ and so $\det(A) = 0$.

It remains to verify the last assertion.

$$\begin{aligned}\det(A) &\equiv \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(k_1,\cdots,k_n) a_{1k_1}\cdots(xa_{k_i} + yb_{k_i})\cdots a_{nk_n}\\ &= x\sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(k_1,\cdots,k_n) a_{1k_1}\cdots a_{k_i}\cdots a_{nk_n}\\ &\quad + y\sum_{(k_1,\cdots,k_n)} \operatorname{sgn}(k_1,\cdots,k_n) a_{1k_1}\cdots b_{k_i}\cdots a_{nk_n}\\ &\equiv x\det(A_1) + y\det(A_2).\end{aligned}$$

The same is true of columns because $\det\left(A^T\right) = \det(A)$ and the rows of A^T are the columns of A.

Definition 6.4.8 *A vector,* $\mathbf{w}$*, is a linear combination of the vectors* $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ *if there exists scalars,* $c_1, \cdots c_r$ *such that* $\mathbf{w} = \sum_{k=1}^{r} c_k \mathbf{v}_k$*. This is the same as saying*

$$\mathbf{w} \in \text{span}\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}.$$

The following corollary is also of great use.

Corollary 6.4.9 *Suppose* A *is an* $n \times n$ *matrix and some column (row) is a linear combination of* r *other columns (rows). Then* $\det(A) = 0$.

Proof: Let $A = \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{pmatrix}$ be the columns of A and suppose the condition that one column is a linear combination of r of the others is satisfied. Then by using Corollary 6.4.7 you may rearrange the columns to have the n^{th} column a linear combination of the first r columns. Thus $\mathbf{a}_n = \sum_{k=1}^{r} c_k \mathbf{a}_k$ and so

$$\det(A) = \det\begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_r & \cdots & \mathbf{a}_{n-1} & \sum_{k=1}^{r} c_k \mathbf{a}_k \end{pmatrix}.$$

By Corollary 6.4.7

$$\det(A) = \sum_{k=1}^{r} c_k \det\begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_r & \cdots & \mathbf{a}_{n-1} & \mathbf{a}_k \end{pmatrix} = 0.$$

The case for rows follows from the fact that $\det(A) = \det\left(A^T\right)$. This proves the corollary.

Recall the following definition of matrix multiplication.

Definition 6.4.10 *If* A *and* B *are* $n \times n$ *matrices,* $A = (a_{ij})$ *and* $B = (b_{ij})$*,* $AB = (c_{ij})$ *where*

$$c_{ij} \equiv \sum_{k=1}^{n} a_{ik} b_{kj}.$$

One of the most important rules about determinants is that the determinant of a product equals the product of the determinants.

Theorem 6.4.11 *Let* A *and* B *be* $n \times n$ *matrices. Then*

$$\det(AB) = \det(A)\det(B).$$

Proof: Let c_{ij} be the ij^{th} entry of AB. Then by Proposition 6.4.4,

$$\det(AB) =$$

$$\begin{aligned}
&\sum_{(k_1,\cdots,k_n)} \text{sgn}(k_1, \cdots, k_n)\, c_{1k_1} \cdots c_{nk_n} \\
= &\sum_{(k_1,\cdots,k_n)} \text{sgn}(k_1, \cdots, k_n) \left(\sum_{r_1} a_{1r_1} b_{r_1 k_1}\right) \cdots \left(\sum_{r_n} a_{nr_n} b_{r_n k_n}\right) \\
= &\sum_{(r_1\cdots,r_n)} \sum_{(k_1,\cdots,k_n)} \text{sgn}(k_1, \cdots, k_n)\, b_{r_1 k_1} \cdots b_{r_n k_n} \left(a_{1r_1} \cdots a_{nr_n}\right) \\
= &\sum_{(r_1\cdots,r_n)} \text{sgn}(r_1 \cdots r_n)\, a_{1r_1} \cdots a_{nr_n} \det(B) = \det(A)\det(B).
\end{aligned}$$

This proves the theorem.

Lemma 6.4.12 *Suppose a matrix is of the form*

$$M = \begin{pmatrix} A & * \\ \mathbf{0} & a \end{pmatrix} \tag{6.16}$$

or

$$M = \begin{pmatrix} A & \mathbf{0} \\ * & a \end{pmatrix} \tag{6.17}$$

where a is a number and A is an $(n-1)\times(n-1)$ matrix and $*$ denotes either a column or a row having length $n-1$ and the $\mathbf{0}$ denotes either a column or a row of length $n-1$ consisting entirely of zeros. Then

$$\det(M) = a\det(A).$$

Proof: Denote M by (m_{ij}). Thus in the first case, $m_{nn} = a$ and $m_{ni} = 0$ if $i \neq n$ while in the second case, $m_{nn} = a$ and $m_{in} = 0$ if $i \neq n$. From the definition of the determinant,

$$\det(M) \equiv \sum_{(k_1,\cdots,k_n)} \operatorname{sgn}_n(k_1,\cdots,k_n)\, m_{1k_1}\cdots m_{nk_n}$$

Letting θ denote the position of n in the ordered list, $(k_1,\cdots,k_n)$ then using the earlier conventions used to prove Lemma 6.4.1, $\det(M)$ equals

$$\sum_{(k_1,\cdots,k_n)} (-1)^{n-\theta} \operatorname{sgn}_{n-1}\left(k_1,\cdots,k_{\theta-1},\overset{\theta}{k_{\theta+1}},\cdots,\overset{n-1}{k_n}\right) m_{1k_1}\cdots m_{nk_n}$$

Now suppose 6.17. Then if $k_n \neq n$, the term involving m_{nk_n} in the above expression equals zero. Therefore, the only terms which survive are those for which $\theta = n$ or in other words, those for which $k_n = n$. Therefore, the above expression reduces to

$$a\sum_{(k_1,\cdots,k_{n-1})} \operatorname{sgn}_{n-1}(k_1,\cdots k_{n-1})\, m_{1k_1}\cdots m_{(n-1)k_{n-1}} = a\det(A).$$

To get the assertion in the situation of 6.16 use Corollary 6.4.6 and 6.17 to write

$$\det(M) = \det\left(M^T\right) = \det\left(\begin{pmatrix} A^T & \mathbf{0} \\ * & a \end{pmatrix}\right) = a\det\left(A^T\right) = a\det(A).$$

This proves the lemma.

In terms of the theory of determinants, arguably the most important idea is that of Laplace expansion along a row or a column. This will follow from the above definition of a determinant.

Definition 6.4.13 *Let $A = (a_{ij})$ be an $n\times n$ matrix. Then a new matrix called the cofactor matrix,* $\operatorname{cof}(A)$ *is defined by* $\operatorname{cof}(A) = (c_{ij})$ *where to obtain c_{ij} delete the i^{th} row and the j^{th} column of A, take the determinant of the $(n-1)\times(n-1)$ matrix which results, (This is called the ij^{th} minor of A.) and then multiply this number by $(-1)^{i+j}$. To make the formulas easier to remember,* $\operatorname{cof}(A)_{ij}$ *will denote the ij^{th} entry of the cofactor matrix.*

The following is the main result. Earlier this was given as a definition and the outrageous totally unjustified assertion was made that the same number would be obtained by expanding the determinant along any row or column. The following theorem proves this assertion.

Theorem 6.4.14 *Let A be an $n \times n$ matrix where $n \geq 2$. Then*

$$\det(A) = \sum_{j=1}^{n} a_{ij} \operatorname{cof}(A)_{ij} = \sum_{i=1}^{n} a_{ij} \operatorname{cof}(A)_{ij}. \tag{6.18}$$

The first formula consists of expanding the determinant along the i^{th} row and the second expands the determinant along the j^{th} column.

Proof: Let $(a_{i1}, \cdots, a_{in})$ be the i^{th} row of A. Let B_j be the matrix obtained from A by leaving every row the same except the i^{th} row which in B_j equals $(0, \cdots, 0, a_{ij}, 0, \cdots, 0)$. Then by Corollary 6.4.7,

$$\det(A) = \sum_{j=1}^{n} \det(B_j)$$

Denote by A^{ij} the $(n-1) \times (n-1)$ matrix obtained by deleting the i^{th} row and the j^{th} column of A. Thus $\operatorname{cof}(A)_{ij} \equiv (-1)^{i+j} \det\left(A^{ij}\right)$. At this point, recall that from Proposition 6.4.4, when two rows or two columns in a matrix, M, are switched, this results in multiplying the determinant of the old matrix by -1 to get the determinant of the new matrix. Therefore, by Lemma 6.4.12,

$$\begin{aligned} \det(B_j) &= (-1)^{n-j}(-1)^{n-i} \det\left(\begin{pmatrix} A^{ij} & * \\ \mathbf{0} & a_{ij} \end{pmatrix}\right) \\ &= (-1)^{i+j} \det\left(\begin{pmatrix} A^{ij} & * \\ \mathbf{0} & a_{ij} \end{pmatrix}\right) = a_{ij} \operatorname{cof}(A)_{ij}. \end{aligned}$$

Therefore,

$$\det(A) = \sum_{j=1}^{n} a_{ij} \operatorname{cof}(A)_{ij}$$

which is the formula for expanding $\det(A)$ along the i^{th} row. Also,

$$\begin{aligned} \det(A) &= \det\left(A^T\right) = \sum_{j=1}^{n} a_{ij}^T \operatorname{cof}\left(A^T\right)_{ij} \\ &= \sum_{j=1}^{n} a_{ji} \operatorname{cof}(A)_{ji} \end{aligned}$$

which is the formula for expanding $\det(A)$ along the i^{th} column. This proves the theorem.

Note that this gives an easy way to write a formula for the inverse of an $n \times n$ matrix.

Theorem 6.4.15 *A^{-1} exists if and only if $\det(A) \neq 0$. If $\det(A) \neq 0$, then $A^{-1} = \left(a_{ij}^{-1}\right)$ where*

$$a_{ij}^{-1} = \det(A)^{-1} \operatorname{cof}(A)_{ji}$$

for $\operatorname{cof}(A)_{ij}$ the ij^{th} cofactor of A.

Proof: By Theorem 6.4.14 and letting $(a_{ir}) = A$, if $\det(A) \neq 0$,

$$\sum_{i=1}^{n} a_{ir} \operatorname{cof}(A)_{ir} \det(A)^{-1} = \det(A) \det(A)^{-1} = 1.$$

Now consider

$$\sum_{i=1}^{n} a_{ir} \operatorname{cof}(A)_{ik} \det(A)^{-1}$$

when $k \neq r$. Replace the k^{th} column with the r^{th} column to obtain a matrix, B_k whose determinant equals zero by Corollary 6.4.7. However, expanding this matrix along the k^{th} column yields

$$0 = \det(B_k) \det(A)^{-1} = \sum_{i=1}^{n} a_{ir} \operatorname{cof}(A)_{ik} \det(A)^{-1}$$

Summarizing,

$$\sum_{i=1}^{n} a_{ir} \operatorname{cof}(A)_{ik} \det(A)^{-1} = \delta_{rk}.$$

Using the other formula in Theorem 6.4.14, and similar reasoning,

$$\sum_{j=1}^{n} a_{rj} \operatorname{cof}(A)_{kj} \det(A)^{-1} = \delta_{rk}$$

This proves that if $\det(A) \neq 0$, then A^{-1} exists with $A^{-1} = \left(a_{ij}^{-1}\right)$, where

$$a_{ij}^{-1} = \operatorname{cof}(A)_{ji} \det(A)^{-1}.$$

Now suppose A^{-1} exists. Then by Theorem 6.4.11,

$$1 = \det(I) = \det\left(AA^{-1}\right) = \det(A) \det\left(A^{-1}\right)$$

so $\det(A) \neq 0$. This proves the theorem.

The next corollary points out that if an $n \times n$ matrix, A has a right or a left inverse, then it has an inverse.

Corollary 6.4.16 *Let A be an $n \times n$ matrix and suppose there exists an $n \times n$ matrix, B such that $BA = I$. Then A^{-1} exists and $A^{-1} = B$. Also, if there exists C an $n \times n$ matrix such that $AC = I$, then A^{-1} exists and $A^{-1} = C$.*

Proof: Since $BA = I$, Theorem 6.4.11 implies

$$\det B \det A = 1$$

and so $\det A \neq 0$. Therefore from Theorem 6.4.15, A^{-1} exists. Therefore,

$$A^{-1} = (BA) A^{-1} = B\left(AA^{-1}\right) = BI = B.$$

The case where $CA = I$ is handled similarly.

The conclusion of this corollary is that left inverses, right inverses and inverses are all the same in the context of $n \times n$ matrices.

Theorem 6.4.15 says that to find the inverse, take the transpose of the cofactor matrix and divide by the determinant. The transpose of the cofactor matrix is called the adjugate or sometimes the classical adjoint of the matrix A. It is an abomination to call it the adjoint although you do sometimes see it referred to in this way. In words, A^{-1} is equal to one over the determinant of A times the adjugate matrix of A.

In case you are solving a system of equations, $A\mathbf{x} = \mathbf{y}$ for $\mathbf{x}$, it follows that if A^{-1} exists,

$$\mathbf{x} = \left(A^{-1}A\right)\mathbf{x} = A^{-1}(A\mathbf{x}) = A^{-1}\mathbf{y}$$

thus solving the system. Now in the case that A^{-1} exists, there is a formula for A^{-1} given above. Using this formula,

$$x_i = \sum_{j=1}^{n} a_{ij}^{-1} y_j = \sum_{j=1}^{n} \frac{1}{\det(A)} \operatorname{cof}(A)_{ji} y_j.$$

By the formula for the expansion of a determinant along a column,

$$x_i = \frac{1}{\det(A)} \det \begin{pmatrix} * & \cdots & y_1 & \cdots & * \\ \vdots & & \vdots & & \vdots \\ * & \cdots & y_n & \cdots & * \end{pmatrix},$$

where here the i^{th} column of A is replaced with the column vector, $(y_1 \cdots, y_n)^T$, and the determinant of this modified matrix is taken and divided by $\det(A)$. This formula is known as Cramer's rule.

Definition 6.4.17 *A matrix M, is upper triangular if $M_{ij} = 0$ whenever $i > j$. Thus such a matrix equals zero below the main diagonal, the entries of the form M_{ii} as shown.*

$$\begin{pmatrix} * & * & \cdots & * \\ 0 & * & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & * \end{pmatrix}$$

A lower triangular matrix is defined similarly as a matrix for which all entries above the main diagonal are equal to zero.

With this definition, here is a simple corollary of Theorem 6.4.14.

Corollary 6.4.18 *Let M be an upper (lower) triangular matrix. Then $\det(M)$ is obtained by taking the product of the entries on the main diagonal.*

6.5 The Cayley Hamilton Theorem*

Definition 6.5.1 *Let A be an $n \times n$ matrix. The characteristic polynomial is defined as*

$$p_A(t) \equiv \det(tI - A)$$

and the solutions to $p_A(t) = 0$ are called eigenvalues. For A a matrix and $p(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$, denote by $p(A)$ the matrix defined by

$$p(A) \equiv A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I.$$

The explanation for the last term is that A^0 is interpreted as I, the identity matrix.

The Cayley Hamilton theorem states that every matrix satisfies its characteristic equation, that equation defined by $P_A(t) = 0$. It is one of the most important theorems in linear algebra[1]. The following lemma will help with its proof.

[1] A special case was first proved by Hamilton in 1853. The general case was announced by Cayley some time later and a proof was given by Frobenius in 1878.

Lemma 6.5.2 *Suppose for all $|\lambda|$ large enough,*

$$A_0 + A_1\lambda + \cdots + A_m\lambda^m = 0,$$

where the A_i are $n \times n$ matrices. Then each $A_i = 0$.

Proof: Multiply by λ^{-m} to obtain

$$A_0\lambda^{-m} + A_1\lambda^{-m+1} + \cdots + A_{m-1}\lambda^{-1} + A_m = 0.$$

Now let $|\lambda| \to \infty$ to obtain $A_m = 0$. With this, multiply by λ to obtain

$$A_0\lambda^{-m+1} + A_1\lambda^{-m+2} + \cdots + A_{m-1} = 0.$$

Now let $|\lambda| \to \infty$ to obtain $A_{m-1} = 0$. Continue multiplying by λ and letting $\lambda \to \infty$ to obtain that all the $A_i = 0$. This proves the lemma.

With the lemma, here is a simple corollary.

Corollary 6.5.3 *Let A_i and B_i be $n \times n$ matrices and suppose*

$$A_0 + A_1\lambda + \cdots + A_m\lambda^m = B_0 + B_1\lambda + \cdots + B_m\lambda^m$$

for all $|\lambda|$ large enough. Then $A_i = B_i$ for all i. Consequently if λ is replaced by any $n \times n$ matrix, the two sides will be equal. That is, for C any $n \times n$ matrix,

$$A_0 + A_1C + \cdots + A_mC^m = B_0 + B_1C + \cdots + B_mC^m.$$

Proof: Subtract and use the result of the lemma.

With this preparation, here is a relatively easy proof of the Cayley Hamilton theorem.

Theorem 6.5.4 *Let A be an $n \times n$ matrix and let $p(\lambda) \equiv \det(\lambda I - A)$ be the characteristic polynomial. Then $p(A) = 0$.*

Proof: Let $C(\lambda)$ equal the transpose of the cofactor matrix of $(\lambda I - A)$ for $|\lambda|$ large. (If $|\lambda|$ is large enough, then λ cannot be in the finite list of eigenvalues of A and so for such λ, $(\lambda I - A)^{-1}$ exists.) Therefore, by Theorem 6.4.15

$$C(\lambda) = p(\lambda)(\lambda I - A)^{-1}.$$

Note that each entry in $C(\lambda)$ is a polynomial in λ having degree no more than $n - 1$. Therefore, collecting the terms,

$$C(\lambda) = C_0 + C_1\lambda + \cdots + C_{n-1}\lambda^{n-1}$$

for C_j some $n \times n$ matrix. It follows that for all $|\lambda|$ large enough,

$$(A - \lambda I)\left(C_0 + C_1\lambda + \cdots + C_{n-1}\lambda^{n-1}\right) = p(\lambda)I$$

and so Corollary 6.5.3 may be used. It follows the matrix coefficients corresponding to equal powers of λ are equal on both sides of this equation. Therefore, if λ is replaced with A, the two sides will be equal. Thus

$$0 = (A - A)\left(C_0 + C_1A + \cdots + C_{n-1}A^{n-1}\right) = p(A)I = p(A).$$

This proves the Cayley Hamilton theorem.

Rank Of A Matrix

7.0.1 Outcomes

A. Recognize and find the row reduced echelon form of a matrix.

B. Determine the rank of a matrix.

C. Describe the row space, column space and null space of a matrix.

D. Define the span of a set of vectors. Recall that a span of vectors in a vector space is a subspace.

E. Determine whether a set of vectors is a subspace.

F. Define linear independence.

G. Determine whether a set of vectors is linearly independent or linearly dependent.

H. Determine a basis and the dimension of a vector space.

I. Characterize the solution set to a matrix equation using rank.

J. Argue that a homogeneous linear system always has a solution and find the solutions.

K. Understand and use the Fredholm alternative.

7.1 Elementary Matrices

The elementary matrices result from doing a row operation to the identity matrix.

Definition 7.1.1 *The row operations consist of the following*

1. *Switch two rows.*
2. *Multiply a row by a nonzero number.*
3. *Replace a row by a multiple of another row added to it.*

The elementary matrices are given in the following definition.

Definition 7.1.2 *The elementary matrices consist of those matrices which result by applying a row operation to an identity matrix. Those which involve switching rows of the identity are called permutation matrices*[1].

[1]More generally, a permutation matrix is a matrix which comes by permuting the rows of the identity matrix, not just switching two rows.

As an example of why these elementary matrices are interesting, consider the following.

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a & b & c & d \\ x & y & z & w \\ f & g & h & i \end{pmatrix} = \begin{pmatrix} x & y & z & w \\ a & b & c & d \\ f & g & h & i \end{pmatrix}$$

A 3×4 matrix was multiplied on the left by an elementary matrix which was obtained from row operation 1 applied to the identity matrix. This resulted in applying the operation 1 to the given matrix. This is what happens in general.

Now consider what these elementary matrices look like. First consider the one which involves switching row i and row j where $i < j$. This matrix is of the form

$$\begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & & & & & & & & & \vdots \\ \vdots & & 1 & & & & & & & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & \cdots & 0 & 1 & \cdots & \cdots & 0 \\ \vdots & & & \vdots & 1 & 0 & \cdots & 0 & & & \vdots \\ \vdots & & & \vdots & & \ddots & & \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 & \cdots & 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & 1 & 0 & \cdots & \cdots & 0 & \cdots & \cdots & 0 \\ \vdots & & & & & & & & 1 & & \vdots \\ \vdots & & & & & & & & & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix}$$

The two exceptional rows are shown. The i^{th} row was the j^{th} and the j^{th} row was the i^{th} in the identity matrix. Now consider what this does to a column vector.

$$\begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & & & & & & & & & \vdots \\ \vdots & & 1 & & & & & & & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & \cdots & 0 & 1 & \cdots & \cdots & 0 \\ \vdots & & & \vdots & 1 & 0 & \cdots & 0 & & & \vdots \\ \vdots & & & \vdots & & \ddots & & \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 & \cdots & 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & 1 & 0 & \cdots & \cdots & 0 & \cdots & \cdots & 0 \\ \vdots & & & & & & & & 1 & & \vdots \\ \vdots & & & & & & & & & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ \vdots \\ v_i \\ \vdots \\ \vdots \\ \vdots \\ v_j \\ \vdots \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} v_1 \\ \vdots \\ \vdots \\ v_j \\ \vdots \\ \vdots \\ \vdots \\ v_i \\ \vdots \\ \vdots \\ v_n \end{pmatrix}$$

Now denote by P^{ij} the elementary matrix which comes from the identity from switching rows i and j. From what was just explained consider multiplication on the left by this

elementary matrix.

$$P^{ij}\begin{pmatrix} a_{11} & a_{12} & \cdots & \cdots & \cdots & \cdots & a_{1p} \\ \vdots & \vdots & & & & & \vdots \\ a_{i1} & a_{i2} & \cdots & \cdots & \cdots & \cdots & a_{ip} \\ \vdots & \vdots & & & & & \vdots \\ a_{j1} & a_{j2} & \cdots & \cdots & \cdots & \cdots & a_{jp} \\ \vdots & \vdots & & & & & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & \cdots & \cdots & a_{np} \end{pmatrix}$$

From the way you multiply matrices this is a matrix which has the indicated columns.

$$\begin{pmatrix} P^{ij}\begin{pmatrix} a_{11} \\ \vdots \\ a_{i1} \\ \vdots \\ a_{j1} \\ \vdots \\ a_{n1} \end{pmatrix}, P^{ij}\begin{pmatrix} a_{12} \\ \vdots \\ a_{i2} \\ \vdots \\ a_{j2} \\ \vdots \\ a_{n2} \end{pmatrix}, \cdots, P^{ij}\begin{pmatrix} a_{1p} \\ \vdots \\ a_{ip} \\ \vdots \\ a_{jp} \\ \vdots \\ a_{np} \end{pmatrix} \end{pmatrix}$$

$$= \left(\begin{pmatrix} a_{11} \\ \vdots \\ a_{j1} \\ \vdots \\ a_{i1} \\ \vdots \\ a_{n1} \end{pmatrix}, \begin{pmatrix} a_{12} \\ \vdots \\ a_{j2} \\ \vdots \\ a_{i2} \\ \vdots \\ a_{n2} \end{pmatrix}, \cdots, \begin{pmatrix} a_{1p} \\ \vdots \\ a_{jp} \\ \vdots \\ a_{ip} \\ \vdots \\ a_{np} \end{pmatrix} \right)$$

$$= \begin{pmatrix} a_{11} & a_{12} & \cdots & \cdots & \cdots & \cdots & a_{1p} \\ \vdots & \vdots & & & & & \vdots \\ a_{j1} & a_{j2} & \cdots & \cdots & \cdots & \cdots & a_{jp} \\ \vdots & \vdots & & & & & \vdots \\ a_{i1} & a_{i2} & \cdots & \cdots & \cdots & \cdots & a_{ip} \\ \vdots & \vdots & & & & & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & \cdots & \cdots & a_{np} \end{pmatrix}$$

This has established the following lemma.

Lemma 7.1.3 *Let P^{ij} denote the elementary matrix which involves switching the i^{th} and the j^{th} rows. Then*

$$P^{ij}A = B$$

where B is obtained from A by switching the i^{th} and the j^{th} rows.

Next consider the row operation which involves multiplying the i^{th} row by a nonzero constant, c. The elementary matrix which results from applying this operation to the i^{th}

row of the identity matrix is of the form

$$\begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & & & & & \vdots \\ \vdots & & 1 & & & & \vdots \\ \vdots & & & c & & & \vdots \\ \vdots & & & & 1 & & \vdots \\ \vdots & & & & & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix}$$

Now consider what this does to a column vector.

$$\begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & & & & & \vdots \\ \vdots & & 1 & & & & \vdots \\ \vdots & & & c & & & \vdots \\ \vdots & & & & 1 & & \vdots \\ \vdots & & & & & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_{i-1} \\ v_i \\ v_{i+1} \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} v_1 \\ \vdots \\ v_{i-1} \\ cv_i \\ v_{i+1} \\ \vdots \\ v_n \end{pmatrix}$$

Denote by $E(c,i)$ this elementary matrix which multiplies the i^{th} row of the identity by the nonzero constant, c. Then from what was just discussed and the way matrices are multiplied,

$$E(c,i) \begin{pmatrix} a_{11} & a_{12} & \cdots & \cdots & \cdots & \cdots & a_{1p} \\ \vdots & \vdots & & & & & \vdots \\ a_{i1} & a_{i2} & \cdots & \cdots & \cdots & \cdots & a_{ip} \\ \vdots & \vdots & & & & & \vdots \\ a_{j2} & a_{j2} & \cdots & \cdots & \cdots & \cdots & a_{jp} \\ \vdots & \vdots & & & & & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & \cdots & \cdots & a_{np} \end{pmatrix}$$

equals a matrix having the columns indicated below.

$$= \left(E(c,i) \begin{pmatrix} a_{11} \\ \vdots \\ a_{i1} \\ \vdots \\ a_{j1} \\ \vdots \\ a_{n1} \end{pmatrix}, E(c,i) \begin{pmatrix} a_{12} \\ \vdots \\ a_{i2} \\ \vdots \\ a_{j2} \\ \vdots \\ a_{n2} \end{pmatrix}, \cdots, E(c,i) \begin{pmatrix} a_{1p} \\ \vdots \\ a_{ip} \\ \vdots \\ a_{jp} \\ \vdots \\ a_{np} \end{pmatrix} \right)$$

$$= \begin{pmatrix} a_{11} & a_{12} & \cdots & \cdots & \cdots & \cdots & a_{1p} \\ \vdots & \vdots & & & & & \vdots \\ ca_{i1} & ca_{i2} & \cdots & \cdots & \cdots & \cdots & ca_{ip} \\ \vdots & \vdots & & & & & \vdots \\ a_{j2} & a_{j2} & \cdots & \cdots & \cdots & \cdots & a_{jp} \\ \vdots & \vdots & & & & & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & \cdots & \cdots & a_{np} \end{pmatrix}$$

This proves the following lemma.

Lemma 7.1.4 *Let $E(c,i)$ denote the elementary matrix corresponding to the row operation in which the i^{th} row is multiplied by the nonzero constant, c. Thus $E(c,i)$ involves multiplying the i^{th} row of the identity matrix by c. Then*

$$E(c,i)\,A = B$$

where B is obtained from A by multiplying the i^{th} row of A by c.

Finally consider the third of these row operations. Denote by $E(c \times i + j)$ the elementary matrix which replaces the j^{th} row with itself added to c times the i^{th} row added to it. In case $i < j$ this will be of the form

$$\begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 & 0 \\ 0 & \ddots & & & & & \vdots \\ \vdots & & 1 & & & & \vdots \\ \vdots & & \vdots & \ddots & & & \vdots \\ \vdots & & c & \cdots & 1 & & \vdots \\ \vdots & & & & & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix}$$

Now consider what this does to a column vector.

$$\begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 & 0 \\ 0 & \ddots & & & & & \vdots \\ \vdots & & 1 & & & & \vdots \\ \vdots & & \vdots & \ddots & & & \vdots \\ \vdots & & c & \cdots & 1 & & \vdots \\ \vdots & & & & & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ v_j \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} v_1 \\ \vdots \\ v_i \\ \vdots \\ cv_i + v_j \\ \vdots \\ v_n \end{pmatrix}$$

Now from this and the way matrices are multiplied,

$$E(c \times i + j) \begin{pmatrix} a_{11} & a_{12} & \cdots & \cdots & \cdots & \cdots & a_{1p} \\ \vdots & \vdots & & & & & \vdots \\ a_{i1} & a_{i2} & \cdots & \cdots & \cdots & \cdots & a_{ip} \\ \vdots & \vdots & & & & & \vdots \\ a_{j2} & a_{j2} & \cdots & \cdots & \cdots & \cdots & a_{jp} \\ \vdots & \vdots & & & & & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & \cdots & \cdots & a_{np} \end{pmatrix}$$

equals a matrix of the following form having the indicated columns.

$$\left(E(c \times i + j) \begin{pmatrix} a_{11} \\ \vdots \\ a_{i1} \\ \vdots \\ a_{j2} \\ \vdots \\ a_{n1} \end{pmatrix}, E(c \times i + j) \begin{pmatrix} a_{12} \\ \vdots \\ a_{i2} \\ \vdots \\ a_{j2} \\ \vdots \\ a_{n2} \end{pmatrix}, \cdots E(c \times i + j) \begin{pmatrix} a_{1p} \\ \vdots \\ a_{ip} \\ \vdots \\ a_{jp} \\ \vdots \\ a_{np} \end{pmatrix} \right)$$

$$= \begin{pmatrix} a_{11} & a_{12} & \cdots & \cdots & \cdots & \cdots & a_{1p} \\ \vdots & \vdots & & & & & \vdots \\ a_{i1} & a_{i2} & \cdots & \cdots & \cdots & \cdots & a_{ip} \\ \vdots & \vdots & & & & & \vdots \\ a_{j2} + ca_{i1} & a_{j2} + ca_{i2} & \cdots & \cdots & \cdots & \cdots & a_{jp} + ca_{ip} \\ \vdots & \vdots & & & & & \vdots \\ a_{n1} & a_{n2} & \cdots & \cdots & \cdots & \cdots & a_{np} \end{pmatrix}$$

The case where $i > j$ is handled similarly. This proves the following lemma.

Lemma 7.1.5 *Let $E(c \times i + j)$ denote the elementary matrix obtained from I by replacing the j^{th} row with c times the i^{th} row added to it. Then*

$$E(c \times i + j) A = B$$

where B is obtained from A by replacing the j^{th} row of A with itself added to c times the i^{th} row of A.

The next theorem is the main result.

Theorem 7.1.6 *To perform any of the three row operations on a matrix, A it suffices to do the row operation on the identity matrix obtaining an elementary matrix, E and then take the product, EA. Furthermore, each elementary matrix is invertible and its inverse is an elementary matrix.*

Proof: The first part of this theorem has been proved in Lemmas 7.1.3 - 7.1.5. It only remains to verify the claim about the inverses. Consider first the elementary matrices corresponding to row operation of type three.

$$E(-c \times i + j) E(c \times i + j) = I$$

This follows because the first matrix takes c times row i in the identity and adds it to row j. When multiplied on the left by $E(-c \times i + j)$ it follows from the first part of this theorem that you take the i^{th} row of $E(c \times i + j)$ which coincides with the i^{th} row of I since that row was not changed, multiply it by $-c$ and add to the j^{th} row of $E(c \times i + j)$ which was the j^{th} row of I added to c times the i^{th} row of I. Thus $E(-c \times i + j)$ multiplied on the left, undoes the row operation which resulted in $E(c \times i + j)$. The same argument applied to the product

$$E(c \times i + j) E(-c \times i + j)$$

replacing c with $-c$ in the argument yields that this product is also equal to I. Therefore, $E(c \times i + j)^{-1} = E(-c \times i + j)$.

Similar reasoning shows that for $E(c, i)$ the elementary matrix which comes from multiplying the i^{th} row by the nonzero constant, c,

$$E(c, i)^{-1} = E\left(c^{-1}, i\right).$$

Finally, consider P^{ij} which involves switching the i^{th} and the j^{th} rows.

$$P^{ij} P^{ij} = I$$

because by the first part of this theorem, multiplying on the left by P^{ij} switches the i^{th} and j^{th} rows of P^{ij} which was obtained from switching the i^{th} and j^{th} rows of the identity. First you switch them to get P^{ij} and then you multiply on the left by P^{ij} which switches these rows again and restores the identity matrix. Thus $\left(P^{ij}\right)^{-1} = P^{ij}$.

7.2 The Row Reduced Echelon Form Of A Matrix

Recall that putting a matrix in row reduced echelon form involves doing row operations as described on Page 33. In this section we review the description of the row reduced echelon form and prove the row reduced echelon form for a given matrix is unique. That is, every matrix can be row reduced to a unique row reduced echelon form. Of course this is not true of the echelon form. The significance of this is that it becomes possible to use the definite article in referring to **the** row reduced echelon form and hence important conclusions about the original matrix may be logically deduced from an examination of its unique row reduced echelon form. First we need the following definition of some terminology.

Definition 7.2.1 *Let $\mathbf{v}_1, \cdots, \mathbf{v}_k, \mathbf{u}$ be vectors. Then $\mathbf{u}$ is said to be a **linear combination** of the vectors $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ if there exist scalars, $c_1, \cdots, c_k$ such that*

$$\mathbf{u} = \sum_{i=1}^{k} c_i \mathbf{v}_i.$$

The collection of all linear combinations of the vectors, $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ *is known as the* ***span*** *of these vectors and is written as* $\text{span}(\mathbf{v}_1, \cdots, \mathbf{v}_k)$.

Another way to say the same thing as expressed in the earlier definition of row reduced echelon form found on Page 32 is the following which is a more useful description when proving the major assertions about the row reduced echelon form.

Definition 7.2.2 *Let* $\mathbf{e}_i$ *denote the column vector which has all zero entries except for the* i^{th} *slot which is one. An* $m \times n$ *matrix is said to be in* ***row reduced echelon form*** *if, in viewing successive columns from left to right, the first nonzero column encountered is* $\mathbf{e}_1$ *and if you have encountered* $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_k$, *the next column is either* $\mathbf{e}_{k+1}$ *or is a linear combination of the vectors,* $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_k$.

Theorem 7.2.3 *Let* A *be an* $m \times n$ *matrix. Then* A *has a row reduced echelon form determined by a simple process.*

Proof: Viewing the columns of A from left to right take the first nonzero column. Pick a nonzero entry in this column and switch the row containing this entry with the top row of A. Now divide this new top row by the value of this nonzero entry to get a 1 in this position and then use row operations to make all entries below this element equal to zero. Thus the first nonzero column is now $\mathbf{e}_1$. Denote the resulting matrix by A_1. Consider the sub-matrix of A_1 to the right of this column and below the first row. Do exactly the same thing for this sub-matrix that was done for A. This time the $\mathbf{e}_1$ will refer to $\mathbb{F}^{m-1}$. Use the first 1 obtained by the above process which is in the top row of this sub-matrix and row operations to zero out every element above it in the rows of A_1. Call the resulting matrix, A_2. Thus A_2 satisfies the conditions of the above definition up to the column just encountered. Continue this way till every column has been dealt with and the result must be in row reduced echelon form.

The following diagram illustrates the above procedure. Say the matrix looked something like the following.

$$\begin{pmatrix} 0 & * & * & * & * & * & * \\ 0 & * & * & * & * & * & * \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & * & * & * & * & * & * \end{pmatrix}$$

First step would yield something like

$$\begin{pmatrix} 0 & 1 & * & * & * & * & * \\ 0 & 0 & * & * & * & * & * \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & * & * & * & * & * \end{pmatrix}$$

For the second step you look at the lower right corner as described,

$$\begin{pmatrix} * & * & * & * & * \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ * & * & * & * & * \end{pmatrix}$$

and if the first column consists of all zeros but the next one is not all zeros, you would get something like this.

$$\begin{pmatrix} 0 & 1 & * & * & * \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & * & * & * \end{pmatrix}$$

Thus, after zeroing out the term in the top row above the 1, you get the following for the next step in the computation of the row reduced echelon form for the original matrix.

$$\begin{pmatrix} 0 & 1 & * & 0 & * & * & * \\ 0 & 0 & 0 & 1 & * & * & * \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & * & * & * \end{pmatrix}.$$

Next you look at the lower right matrix below the top two rows and to the right of the first four columns and repeat the process.

Recall the following definition which was discussed earlier.

Definition 7.2.4 *The first **pivot column** of A is the first nonzero column of A. The next pivot column is the first column after this which becomes $\mathbf{e}_2$ in the row reduced echelon form. The third is the next column which becomes $\mathbf{e}_3$ in the row reduced echelon form and so forth.*

There are three choices for row operations at each step in the above theorem. A natural question is whether the same row reduced echelon matrix always results in the end from following the above algorithm applied in any way. The next corollary says this is the case.

In rough terms, the following lemma states that **linear relationships** between columns in a matrix are preserved by row operations. This simple lemma is the main result in understanding all the major questions related to the row reduced echelon form as well as many other topics.

Lemma 7.2.5 *Let A and B be two $m \times n$ matrices and suppose B results from a row operation applied to A. Then the k^{th} column of B is a linear combination of the $i_1, \cdots, i_r$ columns of B if and only if the k^{th} column of A is a linear combination of the $i_1, \cdots, i_r$ columns of A. Furthermore, the scalars in the linear combination are the same. (The linear relationship between the k^{th} column of A and the $i_1, \cdots, i_r$ columns of A is the same as the linear relationship between the k^{th} column of B and the $i_1, \cdots, i_r$ columns of B.)*

Proof: Let A equal the following matrix in which the $\mathbf{a}_k$ are the columns

$$\begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{pmatrix}$$

and let B equal the following matrix in which the columns are given by the $\mathbf{b}_k$

$$\begin{pmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{pmatrix}$$

Then by Theorem 7.1.6 on Page 111 $\mathbf{b}_k = E\mathbf{a}_k$ where E is an elementary matrix. Suppose then that one of the columns of A is a linear combination of some other columns of A. Say

$$\mathbf{a}_k = \sum_{r \in S} c_r \mathbf{a}_r.$$

Then multiplying by E,

$$\mathbf{b}_k = E\mathbf{a}_k = \sum_{r \in S} c_r E\mathbf{a}_r = \sum_{r \in S} c_r \mathbf{b}_r.$$

This proves the lemma.

Definition 7.2.6 *Two matrices are said to be **row equivalent** if one can be obtained from the other by a sequence of row operations.*

It has been shown above that every matrix is row equivalent to one which is in row reduced echelon form.

Corollary 7.2.7 *The row reduced echelon form is unique. That is if B, C are two matrices in row reduced echelon form and both are row equivalent to A, then $B = C$.*

Proof: Suppose B and C are both row reduced echelon forms for the matrix, A. Then they clearly have the same zero columns since row operations leave zero columns unchanged. If B has the sequence $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_r$ occurring for the first time in the positions, $i_1, i_2, \cdots, i_r$, the description of the row reduced echelon form means that each of these columns is not a linear combination of the preceding columns. Therefore, by Lemma 7.2.5, the same is true of the columns in positions $i_1, i_2, \cdots, i_r$ for C. It follows from the description of the row reduced echelon form that $\mathbf{e}_1, \cdots, \mathbf{e}_r$ occur respectively for the first time in columns $i_1, i_2, \cdots, i_r$ for C. Therefore, both B and C have the sequence $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_r$ occurring for the first time in the positions, $i_1, i_2, \cdots, i_r$. By Lemma 7.2.5, the columns between the i_k and i_{k+1} position in the two matrices are linear combinations involving the same scalars of the columns in the $i_1, \cdots, i_k$ position. Also the columns after the i_r position are linear combinations of the columns in the $i_1, \cdots, i_r$ positions involving the same scalars in both matrices. This is equivalent to the assertion that each of these columns is identical and this proves the corollary.

Now with the above corollary, here is a very fundamental observation.

Corollary 7.2.8 *Suppose A is an $m \times n$ matrix and that $m < n$. That is, the number of rows is less than the number of columns. Then one of the columns of A is a linear combination of the preceding columns of A.*

Proof: Since $m < n$, not all the columns of A can be pivot columns. That is, in the row reduced echelon form say $\mathbf{e}_i$ occurs for the first time at r_i where $r_1 < r_2 < \cdots < r_p$ where $p \leq m$. It follows since $m < n$, there exists some column in the row reduced echelon form which is a linear combination of the preceding columns. By Lemma 7.2.5 the same is true of the columns of A. This proves the corollary.

Example 7.2.9 *Find the row reduced echelon form of the matrix,*

$$\begin{pmatrix} 0 & 0 & 2 & 3 \\ 0 & 2 & 0 & 1 \\ 0 & 1 & 1 & 5 \end{pmatrix}$$

The first nonzero column is the second in the matrix. We switch the third and first rows to obtain

$$\begin{pmatrix} 0 & 1 & 1 & 5 \\ 0 & 2 & 0 & 1 \\ 0 & 0 & 2 & 3 \end{pmatrix}$$

Now we multiply the top row by -2 and add to the second.

$$\begin{pmatrix} 0 & 1 & 1 & 5 \\ 0 & 0 & -2 & -9 \\ 0 & 0 & 2 & 3 \end{pmatrix}$$

Next, add the second row to the bottom and then divide the bottom row by -6

$$\begin{pmatrix} 0 & 1 & 1 & 5 \\ 0 & 0 & -2 & -9 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Next use the bottom row to obtain zeros in the last column above the 1 and divide the second row by -2

$$\begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Finally, add -1 times the middle row to the top.

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

This is in row reduced echelon form.

Example 7.2.10 *Find the row reduced echelon form for the matrix,*

$$\begin{pmatrix} 1 & 2 & 0 & 2 \\ -1 & 3 & 4 & 3 \\ 0 & 5 & 4 & 5 \end{pmatrix}$$

You should verify that the row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & -\frac{8}{5} & 0 \\ 0 & 1 & \frac{4}{5} & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

7.3 The Rank Of A Matrix

7.3.1 The Definition Of Rank

To begin, here is a definition to introduce some terminology.

Definition 7.3.1 *Let A be an $m \times n$ matrix. The **column space** of A is the span of the columns. The **row space** is the span of the rows.*

There are three definitions of the **rank** of a matrix which are useful. These are given in the following definition. It turns out that the concept of **determinant rank** is the one most useful in applications to analysis but is virtually impossible to find directly. The other two concepts of rank are very easily determined and it is a happy fact that all three yield the same number. This is shown later.

Definition 7.3.2 *A **sub-matrix** of a matrix A is a rectangular array of numbers obtained by deleting some rows and columns of A. Let A be an $m \times n$ matrix. The **determinant rank** of the matrix equals r where r is the largest number such that some $r \times r$ sub-matrix of A has a non zero determinant. The **row space** of a matrix is the span of the rows and the **column space** of a matrix is the span of the columns. The **row rank** of a matrix is the number of nonzero rows in the row reduced echelon form and the **column rank** is the number columns in the row reduced echelon form which are one of the $\mathbf{e}_k$ vectors. Thus the column rank equals the number of pivot columns. It follows the row rank equals the column rank. This is also called the rank of the matrix. The rank of a matrix, A is denoted by* $\operatorname{rank}(A)$.

Example 7.3.3 *Consider the matrix,*

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{pmatrix}$$

What is its rank?

You could look at all the 2×2 submatrices

$$\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}, \begin{pmatrix} 1 & 3 \\ 2 & 6 \end{pmatrix}, \begin{pmatrix} 2 & 3 \\ 4 & 6 \end{pmatrix}.$$

Each has determinant equal to 0. Therefore, the rank is less than 2. Now look at the 1×1 submatrices. There exists one of these which has nonzero determinant. For example (1) has determinant equal to 1 and so the rank of this matrix equals 1.

Of course this example was pretty easy but what if you had a 4×7 matrix? You would have to consider all the 4×4 submatrices and then all the 3×3 submatrices and then all the 2×2 matrices and finally all the 1×1 matrices in order to compute the rank. Clearly this is not practical. The following theorem will remove the difficulties just indicated.

The following theorem is proved later.

Theorem 7.3.4 *Let A be an $m \times n$ matrix. Then the row rank, column rank and determinant rank are all the same.*

Example 7.3.5 *Find the rank of the matrix,*

$$\begin{pmatrix} 1 & 2 & 1 & 3 & 0 \\ -4 & 3 & 2 & 1 & 2 \\ 3 & 2 & 1 & 6 & 5 \\ 4 & -3 & -2 & 1 & 7 \end{pmatrix}.$$

From the above definition, all you have to do is find the row reduced echelon form and then count up the number of nonzero rows. But the row reduced echelon form of this matrix is

$$\begin{pmatrix} 1 & 0 & 0 & 0 & -\frac{17}{4} \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & -\frac{45}{4} \\ 0 & 0 & 0 & 1 & \frac{9}{2} \end{pmatrix}$$

and so the rank of this matrix is 4.

Find the rank of the matrix

$$\begin{pmatrix} 1 & 2 & 1 & 3 & 0 \\ -4 & 3 & 2 & 1 & 2 \\ 3 & 2 & 1 & 6 & 5 \\ 0 & 7 & 4 & 10 & 7 \end{pmatrix}$$

The row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & 0 & \frac{3}{2} & \frac{5}{2} \\ 0 & 1 & 0 & -4 & -17 \\ 0 & 0 & 1 & \frac{19}{2} & \frac{63}{2} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and so this time the rank is 3.

7.3.2 Finding The Row And Column Space Of A Matrix

The row reduced echelon form also can be used to obtain an efficient description of the row and column space of a matrix. Of course you can get the column space by simply saying that it equals the span of all the columns but often you can get the column space as the span of fewer columns than this. This is what we mean by an "efficient description". This is illustrated in the next example.

Example 7.3.6 *Find the rank of the following matrix and describe the column and row spaces efficiently.*

$$\begin{pmatrix} 1 & 2 & 1 & 3 & 2 \\ 1 & 3 & 6 & 0 & 2 \\ 3 & 7 & 8 & 6 & 6 \end{pmatrix} \tag{7.1}$$

The row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & -9 & 9 & 2 \\ 0 & 1 & 5 & -3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Therefore, the rank of this matrix equals 2. All columns of this row reduced echelon form are in

$$\text{span}\left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\right).$$

For example,

$$\begin{pmatrix} -9 \\ 5 \\ 0 \end{pmatrix} = -9\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 5\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

By Lemma 7.2.5, all columns of the original matrix, are similarly contained in the span of the first two columns of that matrix. For example, consider the third column of the original matrix.

$$\begin{pmatrix} 1 \\ 6 \\ 8 \end{pmatrix} = -9\begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} + 5\begin{pmatrix} 2 \\ 3 \\ 7 \end{pmatrix}.$$

How did I know to use -9 and 5 for the coefficients? This is what Lemma 7.2.5 says! It says linear relationships are all preserved. Therefore, the column space of the original matrix equals the span of the first two columns. This is the desired efficient description of the column space.

What about an efficient description of the row space? When row operations are used, the resulting vectors remain in the row space. Thus the rows in the row reduced echelon form are in the row space of the original matrix. Furthermore, by reversing the row operations, each row of the original matrix can be obtained as a linear combination of the rows in the row reduced echelon form. It follows that the span of the nonzero rows in the row reduced echelon equals the span of the original rows. In the above example, the row space equals the span of the two vectors, $\begin{pmatrix} 1 & 0 & -9 & 9 & 2 \end{pmatrix}$ and $\begin{pmatrix} 0 & 1 & 5 & -3 & 0 \end{pmatrix}$.

Example 7.3.7 *Find the rank of the following matrix and describe the column and row spaces efficiently.*

$$\begin{pmatrix} 1 & 2 & 1 & 3 & 2 \\ 1 & 3 & 6 & 0 & 2 \\ 1 & 2 & 1 & 3 & 2 \\ 1 & 3 & 2 & 4 & 0 \end{pmatrix} \tag{7.2}$$

The row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \frac{13}{2} \\ 0 & 1 & 0 & 2 & -\frac{5}{2} \\ 0 & 0 & 1 & -1 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

and so the rank is 3, the row space is the span of the vectors,

$$\begin{pmatrix} 0 & 0 & 1 & -1 & \frac{1}{2} \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 & 2 & -\frac{5}{2} \end{pmatrix},$$
$$\begin{pmatrix} 1 & 0 & 0 & 0 & \frac{13}{2} \end{pmatrix},$$

and the column space is the span of the first three columns in the **original matrix**,

$$\text{span}\left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 1 \\ 6 \\ 1 \\ 2 \end{pmatrix}\right).$$

Example 7.3.8 *Find the rank of the following matrix and describe the column and row spaces efficiently.*

$$\begin{pmatrix} 1 & 2 & 3 & 0 & 1 \\ 2 & 1 & 3 & 2 & 4 \\ -1 & 2 & 1 & 3 & 1 \end{pmatrix}.$$

The row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & 1 & 0 & \frac{21}{17} \\ 0 & 1 & 1 & 0 & -\frac{2}{17} \\ 0 & 0 & 0 & 1 & \frac{14}{17} \end{pmatrix}.$$

It follows the rank is three and the column space is the span of the first, second and fourth columns of the **original matrix.**

$$\text{span}\left(\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 3 \end{pmatrix}\right)$$

while the row space is the span of the vectors $\begin{pmatrix} 0 & 0 & 0 & 1 & \frac{14}{17} \end{pmatrix}, \begin{pmatrix} 0 & 1 & 1 & 0 & -\frac{2}{17} \end{pmatrix}$, and $\begin{pmatrix} 1 & 0 & 1 & 0 & \frac{21}{17} \end{pmatrix}$.

Procedure 7.3.9 *To find the rank of a matrix, obtain the row reduced echelon form for the matrix. Then count the number of nonzero rows or equivalently the number of pivot columns. This is the rank. The row space is the span of the nonzero rows in the row reduced echelon form and the column space is the span of the pivot columns of the* ***original matrix****.*

7.4 Linear Independence And Bases

7.4.1 Linear Independence And Dependence

First we consider the concept of linear independence. We define what it means for vectors in $\mathbb{F}^n$ to be linearly independent and then give equivalent descriptions. In the following definition, the symbol,

$$\begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \end{pmatrix}$$

denotes the matrix which has the vector, $\mathbf{v}_1$ as the first column, $\mathbf{v}_2$ as the second column and so forth until $\mathbf{v}_k$ is the k^{th} column.

Definition 7.4.1 *Let $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ be vectors in $\mathbb{F}^n$. Then this collection of vectors is said to be **linearly independent** if for each $r \geq 0$, each of the first k columns of the $n \times (k+r)$ matrix $\begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k & \mathbf{w}_1 & \cdots & \mathbf{w}_r \end{pmatrix}$ is a pivot column. Thus the first k columns in the row reduced echelon form are $\mathbf{e}_1, \cdots, \mathbf{e}_k$.*

Here is what the above means in terms of linear relationships.

Corollary 7.4.2 *The collection of vectors, $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is linearly independent if and only if none of these vectors is a linear combination of the others.*

Proof: If $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is linearly independent, then every column in

$$\begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \end{pmatrix}$$

is a pivot column which requires that the row reduced echelon form is

$$\begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_k \end{pmatrix}.$$

Now none of the $\mathbf{e}_i$ vectors is a linear combination of the others. By Lemma 7.2.5 on Page 113 none of the $\mathbf{v}_i$ is a linear combination of the others. Recall this lemma says linear relationships between the columns are preserved under row operations.

Next suppose none of the columns in $\begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \end{pmatrix}$ is a linear combination of the others and consider a matrix of the form

$$\begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k & \mathbf{w}_1 & \cdots & \mathbf{w}_r \end{pmatrix}. \tag{7.3}$$

Then by Lemma 7.2.5 the same is true of the first k columns of the row reduced echelon form for this matrix. From the description of the row reduced echelon form, it follows that for $i \leq k$, the i^{th} column of the row reduced echelon form must be $\mathbf{e}_i$ since otherwise, it would either equal the zero vector in which case it would be a linear combination of the first k vectors or else it would be a linear combination of the vectors, $\mathbf{e}_1, \cdots, \mathbf{e}_{i-1}$ and either situation would require the i^{th} vector to be a linear combination of the other vectors in the first k columns in the row reduced echelon form. Therefore, by Lemma 7.2.5, the same linear relation would exist for the first k columns of 7.3. Therefore, each of the first k columns in column in $\begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k & \mathbf{w}_1 & \cdots & \mathbf{w}_r \end{pmatrix}$ is a pivot column.

Corollary 7.4.3 *The collection of vectors, $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is linearly independent if and only if whenever*

$$\sum_{i=1}^{n} c_i \mathbf{v}_i = \mathbf{0}$$

it follows each $c_i = 0$.

Proof: Suppose first $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is linearly independent. Then by Corollary 7.4.2, none of the vectors is a linear combination of the others. Now suppose

$$\sum_{i=1}^{n} c_i \mathbf{v}_i = \mathbf{0}$$

and not all the $c_i = 0$. Then pick c_i which is not zero, divide by it and solve for $\mathbf{v}_i$ in terms of the other $\mathbf{v}_j$, contradicting the fact that none of the $\mathbf{v}_i$ equals a linear combination of the others.

Now suppose the condition about the sum holds. If $\mathbf{v}_i$ is a linear combination of the other vectors in the list, then you could obtain an equation of the form

$$\mathbf{v}_i = \sum_{j \neq i} c_j \mathbf{v}_j$$

and so

$$\mathbf{0} = \sum_{j \neq i} c_j \mathbf{v}_j + (-1)\, \mathbf{v}_i,$$

contradicting the condition about the sum.

Sometimes we refer to this last condition about sums as follows: The set of vectors, $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is linearly independent if and only if there is no nontrivial linear combination which equals zero. (A nontrivial linear combination is one in which not all the scalars equal zero.)

We give the following equivalent definition of linear independence which follows from the above corollaries.

Definition 7.4.4 *A set of vectors, $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is linearly independent if none of the vectors is a linear combination of the others or equivalently if there is no nontrivial linear combination of the vectors which equals 0. It is said to be* ***linearly dependent*** *if at least one of the vectors* ***is*** *a linear combination of the others or equivalently there exists a nontrivial linear combination which equals zero.*

Note the meaning of the words. To say a set of vectors is linearly dependent means at least one is a linear combination of the others. In other words, it is in a sense "dependent" on these other vectors.

The following corollary follows right away from the row reduced echelon form.

Corollary 7.4.5 *Let $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ be a set of vectors in $\mathbb{F}^n$. Then if $k > n$, it must be the case that $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is not linearly independent. In other words, if $k > n$, then $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is dependent.*

Proof: If $k > n$, then the columns of $\begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \end{pmatrix}$ cannot each be a pivot column because there are at most n pivot columns due to the fact the matrix has only n rows.

It follows from Corollary 7.4.2 that if $k > n$, then at least one of the vectors in $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is a linear combination of the others.

Example 7.4.6 *Determine whether the vectors, $\left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \\ 0 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 2 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ 2 \\ -1 \end{pmatrix} \right\}$ are linearly independent. If they are linearly dependent, exhibit one of the vectors as a linear combination of the others.*

Form the matrix mentioned above.

$$\begin{pmatrix} 1 & 2 & 0 & 3 \\ 2 & 1 & 1 & 2 \\ 3 & 0 & 1 & 2 \\ 0 & 1 & 2 & -1 \end{pmatrix}$$

Then the row reduced echelon form of this matrix is

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Thus not all the columns are pivot columns and so the vectors are not linear independent. Note the fourth column is of the form

$$1\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + 1\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + (-1)\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

From Lemma 7.2.5, the same linear relationship exists between the columns of the original matrix. Thus

$$1\begin{pmatrix} 1 \\ 2 \\ 3 \\ 0 \end{pmatrix} + 1\begin{pmatrix} 2 \\ 1 \\ 0 \\ 1 \end{pmatrix} + (-1)\begin{pmatrix} 0 \\ 1 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 2 \\ -1 \end{pmatrix}.$$

Note the usefulness of the row reduced echelon form in discovering hidden linear relationships in collections of vectors.

Example 7.4.7 *Determine whether the vectors,* $\left\{\begin{pmatrix} 1 \\ 2 \\ 3 \\ 0 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \\ 0 \\ 1 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \\ 1 \\ 2 \end{pmatrix}\begin{pmatrix} 3 \\ 2 \\ 2 \\ 0 \end{pmatrix}\right\}$ *are linearly independent. If they are linearly dependent, exhibit one of the vectors as a linear combination of the others.*

The matrix used to find this is

$$\begin{pmatrix} 1 & 2 & 0 & 3 \\ 2 & 1 & 1 & 2 \\ 3 & 0 & 1 & 2 \\ 0 & 1 & 2 & 0 \end{pmatrix}$$

The row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and so every column is a pivot column. Therefore, these vectors are linearly independent and there is no way to obtain one of the vectors as a linear combination of the others.

7.4.2 Subspaces

It turns out that the span of a set of vectors is something called a subspace. We will now give a different, easier to remember description of subspaces and will then show that every subspace is the span of a set of vectors.

Definition 7.4.8 *Let V be a nonempty collection of vectors in $\mathbb{F}^n$. Then V is called a subspace if whenever α, β are scalars and $\mathbf{u}, \mathbf{v}$ are vectors in V, the linear combination, $\alpha\mathbf{u} + \beta\mathbf{v}$ is also in V.*

Theorem 7.4.9 *V is a subspace of $\mathbb{F}^n$ if and only if there exist vectors of V, $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ such that $V = \operatorname{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k)$.*

Proof: Pick a vector of $V, \mathbf{u}_1$. If $V = \operatorname{span}\{\mathbf{u}_1\}$, then stop. You have found your list of vectors. If $V \neq \operatorname{span}(\mathbf{u}_1)$, then there exists $\mathbf{u}_2$ a vector of V which is not a vector in $\operatorname{span}(\mathbf{u}_1)$. Consider $\operatorname{span}(\mathbf{u}_1, \mathbf{u}_2)$. If $V = \operatorname{span}(\mathbf{u}_1, \mathbf{u}_2)$, stop. Otherwise, pick $\mathbf{u}_3 \notin \operatorname{span}(\mathbf{u}_1, \mathbf{u}_2)$. Continue this way. The process must stop with $\mathbf{u}_k$ for some $k \leq n$ since otherwise, the matrix

$$\begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_k \end{pmatrix}$$

having these vectors as columns would have n rows and $k > n$ columns. Consequently, it can have no more than n pivot columns and so the first column which is not a pivot column would be a linear combination of the preceding columns contrary to the construction.

For the other half, suppose $V = \operatorname{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k)$ and let $\sum_{i=1}^k c_i\mathbf{u}_i$ and $\sum_{i=1}^k d_i\mathbf{u}_i$ be two vectors in V. Now let α and β be two scalars. Then

$$\alpha\sum_{i=1}^k c_i\mathbf{u}_i + \beta\sum_{i=1}^k d_i\mathbf{u}_i = \sum_{i=1}^k (\alpha c_i + \beta d_i)\,\mathbf{u}_i$$

which is one of the things in $\operatorname{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k)$ showing that $\operatorname{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k)$ has the properties of a subspace. This proves the theorem.

The following corollary also follows easily.

Corollary 7.4.10 *If V is a subspace of $\mathbb{F}^n$, then there exist vectors of $V, \{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ such that $V = \operatorname{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k)$ and $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ is linearly independent.*

Proof: Let $V = \operatorname{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k)$. Then let the vectors $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ be the columns of the following matrix.

$$\begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_k \end{pmatrix}$$

Retain only the pivot columns. That is, determine the pivot columns from the row reduced echelon form and these are a basis for $\operatorname{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k)$.

The message is that subspaces of $\mathbb{F}^n$ consist of spans of finite, linearly independent collections of vectors of $\mathbb{F}^n$.

7.4.3 Basis Of A Subspace

It was just shown in Corollary 7.4.10 that every subspace of $\mathbb{F}^n$ is equal to the span of a linearly independent collection of vectors of $\mathbb{F}^n$. Such a collection of vectors is called a basis.

Definition 7.4.11 *Let V be a subspace of $\mathbb{F}^n$. Then $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ is a **basis** for V if the following two conditions hold.*

1. $\operatorname{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k) = V$.
2. *$\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ is linearly independent.*

*The plural of basis is **bases**.*

The main theorem about bases is the following.

Theorem 7.4.12 *Let V be a subspace of $\mathbb{F}^n$ and suppose $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ and $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ are two bases for V. Then $k = m$.*

Proof: Suppose $k < m$. Then since $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ is a basis for V, each $\mathbf{v}_i$ is a linear combination of the vectors of $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$. Consider the matrix

$$\begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_k & \mathbf{v}_1 & \cdots & \mathbf{v}_m \end{pmatrix}$$

in which each of the $\mathbf{u}_i$ is a pivot column because the $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ are linearly independent. Therefore, the row reduced echelon form of this matrix is

$$\begin{pmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_k & \mathbf{w}_1 & \cdots & \mathbf{w}_m \end{pmatrix} \tag{7.4}$$

where each $\mathbf{w}_j$ has zeroes below the k^{th} row. This is because of Lemma 7.2.5 which implies each $\mathbf{w}_i$ is a linear combination of the $\mathbf{e}_1, \cdots, \mathbf{e}_k$. Discarding the bottom $n-k$ rows of zeroes in the above, yields the matrix,

$$\begin{pmatrix} \mathbf{e}'_1 & \cdots & \mathbf{e}'_k & \mathbf{w}'_1 & \cdots & \mathbf{w}'_m \end{pmatrix}$$

in which all vectors are in $\mathbb{F}^k$. Since $m > k$, it follows from Corollary 7.4.5 that the vectors, $\{\mathbf{w}'_1, \cdots, \mathbf{w}'_m\}$ are dependent. Therefore, some $\mathbf{w}'_j$ is a linear combination of the other $\mathbf{w}'_i$. Therefore, $\mathbf{w}_j$ is a linear combination of the other $\mathbf{w}_i$ in 7.4. By Lemma 7.2.5 again, the same linear relationship exists between the $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ showing that $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ is not linearly independent and contradicting the assumption that $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ is a basis. It follows $m \leq k$. Similarly, $k \leq m$. This proves the theorem.

This is a very important theorem so here is another proof of it.

Theorem 7.4.13 *Let V be a subspace and suppose $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ and $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ are two bases for V. Then $k = m$.*

Proof: Suppose $k > m$. Then since the vectors, $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ span V, there exist scalars, c_{ij} such that

$$\sum_{i=1}^{m} c_{ij}\mathbf{v}_i = \mathbf{u}_j.$$

Therefore,

$$\sum_{j=1}^{k} d_j\mathbf{u}_j = \mathbf{0} \text{ if and only if } \sum_{j=1}^{k}\sum_{i=1}^{m} c_{ij}d_j\mathbf{v}_i = \mathbf{0}$$

if and only if

$$\sum_{i=1}^{m}\left(\sum_{j=1}^{k} c_{ij}d_j\right)\mathbf{v}_i = \mathbf{0}$$

Now since$\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is independent, this happens if and only if

$$\sum_{j=1}^{k} c_{ij}d_j = 0,\ i = 1, 2, \cdots, m.$$

However, this is a system of m equations in k variables, $d_1, \cdots, d_k$ and $m < k$. Therefore, there exists a solution to this system of equations in which not all the d_j are equal to zero. Recall why this is so. The augmented matrix for the system is of the form $\begin{pmatrix} C & \mathbf{0} \end{pmatrix}$ where C is a matrix which has more columns than rows. Therefore, there are free variables and hence nonzero solutions to the system of equations. However, this contradicts the linear independence of $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ because, as explained above, $\sum_{j=1}^{k} d_j\mathbf{u}_j = \mathbf{0}$. Similarly it cannot happen that $m > k$. This proves the theorem.

The following definition can now be stated.

Definition 7.4.14 *Let V be a subspace of $\mathbb{F}^n$. Then the **dimension** of V is defined to be the number of vectors in a basis.*

Corollary 7.4.15 *The dimension of $\mathbb{F}^n$ is n.*

Proof: You only need to exhibit a basis for $\mathbb{F}^n$ which has n vectors. Such a basis is $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$.

Corollary 7.4.16 *Suppose $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is linearly independent and each $\mathbf{v}_i$ is a vector in $\mathbb{F}^n$. Then $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is a basis for $\mathbb{F}^n$. Suppose $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ spans $\mathbb{F}^n$. Then $m \geq n$. If $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ spans $\mathbb{F}^n$, then $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is linearly independent.*

Proof: Let $\mathbf{u}$ be a vector of $\mathbb{F}^n$ and consider the matrix,

$$\begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n & \mathbf{u} \end{pmatrix}.$$

Since each $\mathbf{v}_i$ is a pivot column, the row reduced echelon form is

$$\begin{pmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_n & \mathbf{w} \end{pmatrix}$$

and so, since $\mathbf{w}$ is in $\text{span}(\mathbf{e}_1, \cdots, \mathbf{e}_n)$, it follows from Lemma 7.2.5 that $\mathbf{u}$ is one of the vectors in $\text{span}(\mathbf{v}_1, \cdots, \mathbf{v}_n)$. Therefore, $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is a basis as claimed.

To establish the second claim, suppose that $m < n$. Then letting $\mathbf{v}_{i_1}, \cdots, \mathbf{v}_{i_k}$ be the pivot columns of the matrix

$$\begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_m \end{pmatrix}$$

it follows $k \leq m < n$ and these k pivot columns would be a basis for $\mathbb{F}^n$ having fewer than n vectors, contrary to Theorem 7.4.12 which states every two bases have the same number of vectors in them.

Finally consider the third claim. If $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is not linearly independent, then replace this list with $\{\mathbf{v}_{i_1}, \cdots, \mathbf{v}_{i_k}\}$ where these are the pivot columns of the matrix,

$$\begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{pmatrix}$$

Then $\{\mathbf{v}_{i_1}, \cdots, \mathbf{v}_{i_k}\}$ spans $\mathbb{F}^n$ and is linearly independent so it is a basis having less than n vectors contrary to Theorem 7.4.12 which states every two bases have the same number of vectors in them. This proves the corollary.

Example 7.4.17 *Find the rank of the following matrix. If the rank is r, identify r columns **in the original matrix** which have the property that every other column may be written as a linear combination of these. Also find a basis for the row and column spaces of the matrices.*

$$\begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & 5 & -4 & -1 \\ -2 & 3 & 1 & 0 \end{pmatrix}$$

The row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & 0 & \frac{27}{70} \\ 0 & 1 & 0 & \frac{1}{10} \\ 0 & 0 & 1 & \frac{33}{70} \end{pmatrix}$$

and so the rank of the matrix is 3. A basis for the column space is the first three columns of the **original matrix.** I know they span because the first three columns of the row reduced echelon form above span the column space of that matrix. They are linearly independent because the first three columns of the row reduced echelon form are linearly independent. By Lemma 7.2.5 all linear relationships are preserved and so these first three vectors form a basis for the column space. The four rows of the row reduced echelon form form a basis for the row space of the original matrix.

Example 7.4.18 *Find the rank of the following matrix. If the rank is r, identify r columns* ***in the original matrix*** *which have the property that every other column may be written as a linear combination of these. Also find a basis for the row and column spaces of the matrices.*

$$\begin{pmatrix} 1 & 2 & 3 & 0 & 1 \\ 1 & 1 & 2 & -6 & 2 \\ -2 & 3 & 1 & 0 & 2 \end{pmatrix}$$

The row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & 1 & 0 & -\frac{1}{7} \\ 0 & 1 & 1 & 0 & \frac{4}{7} \\ 0 & 0 & 0 & 1 & -\frac{11}{42} \end{pmatrix}.$$

A basis for the column space of this row reduced echelon form is the first second and fourth columns. Therefore, a basis for the column space in the **original matrix** is the first second and fourth columns. The rank of the matrix is 3. A basis for the row space of the original matrix is the columns of the row reduced echelon form.

7.4.4 Extending An Independent Set To Form A Basis

Suppose $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ is a linearly independent set of vectors in $\mathbb{F}^n$. It turns out there is a larger set of vectors, $\{\mathbf{v}_1, \cdots, \mathbf{v}_m, \mathbf{v}_{m+1}, \cdots, \mathbf{v}_n\}$ which is a basis for $\mathbb{F}^n$. It is easy to do this using the row reduced echelon form. Consider the following matrix having rank n in which the columns are shown.

$$\begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_m & \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{pmatrix}.$$

Since the $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ are linearly independent, the row reduced echelon form of this matrix is of the form

$$\begin{pmatrix} \mathbf{e}_1 & \cdots & \mathbf{e}_m & \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$$

Now the pivot columns can be identified and this leads to a basis for the column space of the original matrix which is of the form

$$\left\{\mathbf{v}_1, \cdots, \mathbf{v}_m, \mathbf{e}_{i_1}, \cdots, \mathbf{e}_{i_{n-m}}\right\}.$$

This proves the following theorem.

Theorem 7.4.19 *Let $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ be a linearly independent set of vectors in $\mathbb{F}^n$. Then there is a larger set of vectors, $\{\mathbf{v}_1, \cdots, \mathbf{v}_m, \mathbf{v}_{m+1}, \cdots, \mathbf{v}_n\}$ which is a basis for $\mathbb{F}^n$.*

Example 7.4.20 *The vectors,* $\left\{\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}\right\}$ *are linearly independent. Enlarge this set of vectors to form a basis for $\mathbb{R}^4$.*

Using the above technique, consider the following matrix.

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

whose row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

The pivot columns are numbers 1,2,3, and 6. Therefore, a basis is

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

7.4.5 Finding The Null Space Or Kernel Of A Matrix

Let A be an $m \times n$ matrix.

Definition 7.4.21 $\ker(A)$*, also referred to as the null space of* A *is defined as follows.*

$$\ker(A) = \{\mathbf{x} : A\mathbf{x} = \mathbf{0}\}$$

and to find $\ker(A)$ *one must solve the system of equations* $A\mathbf{x} = \mathbf{0}$.

This is not new! There is just some new terminology being used. To repeat, $\ker(A)$ is the solution to the system $A\mathbf{x} = \mathbf{0}$.

Example 7.4.22 *Let*

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & -1 & 1 \\ 2 & 3 & 3 \end{pmatrix}.$$

Find $\ker(A)$.

You need to solve the equation $A\mathbf{x} = \mathbf{0}$. To do this you write the augmented matrix and then obtain the row reduced echelon form and the solution. The augmented matrix is

$$\left(\begin{array}{ccc|c} 1 & 2 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 2 & 3 & 3 & 0 \end{array}\right)$$

Next place this matrix in row reduced echelon form,

$$\left(\begin{array}{ccc|c} 1 & 0 & 3 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Note that x_1 and x_2 are basic variables while x_3 is a free variable. Therefore, the solution to this system of equations, $A\mathbf{x} = \mathbf{0}$ is given by

$$\begin{pmatrix} 3t \\ t \\ t \end{pmatrix} : t \in \mathbb{R}.$$

Example 7.4.23 *Let*

$$A = \begin{pmatrix} 1 & 2 & 1 & 0 & 1 \\ 2 & -1 & 1 & 3 & 0 \\ 3 & 1 & 2 & 3 & 1 \\ 4 & -2 & 2 & 6 & 0 \end{pmatrix}$$

Find the null space of A.

You need to solve the equation, $A\mathbf{x} = \mathbf{0}$. The augmented matrix is

$$\left(\begin{array}{ccccc|c} 1 & 2 & 1 & 0 & 1 & 0 \\ 2 & -1 & 1 & 3 & 0 & 0 \\ 3 & 1 & 2 & 3 & 1 & 0 \\ 4 & -2 & 2 & 6 & 0 & 0 \end{array}\right)$$

Its row reduced echelon form is

$$\left(\begin{array}{ccccc|c} 1 & 0 & \frac{3}{5} & \frac{6}{5} & \frac{1}{5} & 0 \\ 0 & 1 & \frac{1}{5} & -\frac{3}{5} & \frac{2}{5} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

It follows x_1 and x_2 are basic variables and x_3, x_4, x_5 are free variables. Therefore, $\ker(A)$ is given by

$$\left\{ \begin{pmatrix} \left(-\frac{3}{5}\right) s_1 + \left(\frac{-6}{5}\right) s_2 + \left(\frac{1}{5}\right) s_3 \\ \left(-\frac{1}{5}\right) s_1 + \left(\frac{3}{5}\right) s_2 + \left(-\frac{2}{5}\right) s_3 \\ s_1 \\ s_2 \\ s_3 \end{pmatrix} : s_1, s_2, s_3 \in \mathbb{R}. \right.$$

We write this in the form

$$s_1 \begin{pmatrix} -\frac{3}{5} \\ -\frac{1}{5} \\ 1 \\ 0 \\ 0 \end{pmatrix} + s_2 \begin{pmatrix} \frac{-6}{5} \\ \frac{3}{5} \\ 0 \\ 1 \\ 0 \end{pmatrix} + s_3 \begin{pmatrix} \frac{1}{5} \\ -\frac{2}{5} \\ 0 \\ 0 \\ 1 \end{pmatrix} : s_1, s_2, s_3 \in \mathbb{R}.$$

In other words, the null space of this matrix equals the span of the three vectors above. Thus

$$\ker(A) = \text{span}\left(\begin{pmatrix} -\frac{3}{5} \\ -\frac{1}{5} \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{-6}{5} \\ \frac{3}{5} \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{5} \\ -\frac{2}{5} \\ 0 \\ 0 \\ 1 \end{pmatrix} \right).$$

This is the same as

$$\ker(A) = \text{span}\left(\begin{pmatrix} \frac{3}{5} \\ \frac{1}{5} \\ -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{6}{5} \\ \frac{-3}{5} \\ 0 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{-1}{5} \\ \frac{2}{5} \\ 0 \\ 0 \\ -1 \end{pmatrix} \right).$$

Notice also that the three vectors above are linearly independent and so the dimension of $\ker(A)$ is 3. This is generally the way it works. The number of free variables equals the dimension of the null space while the number of basic variables equals the number of pivot columns which equals the rank. We state this in the following theorem.

Definition 7.4.24 *The dimension of the null space of a matrix is called the* ***nullity***[2] *and written as* $\text{null}(A)$.

Theorem 7.4.25 *Let A be an $m \times n$ matrix. Then* $\text{rank}(A) + \text{null}(A) = n$.

7.4.6 Rank And Existence Of Solutions To Linear Systems

Consider the linear system of equations,

$$A\mathbf{x} = \mathbf{b} \tag{7.5}$$

where A is an $m \times n$ matrix, $\mathbf{x}$ is a $n \times 1$ column vector, and $\mathbf{b}$ is an $m \times 1$ column vector. Suppose

$$A = \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{pmatrix}$$

where the $\mathbf{a}_k$ denote the columns of A. Then $\mathbf{x} = (x_1, \cdots, x_n)^T$ is a solution of the system 7.5, if and only if

$$x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$$

which says that $\mathbf{b}$ is a vector in $\text{span}(\mathbf{a}_1, \cdots, \mathbf{a}_n)$. This shows that there exists a solution to the system, 7.5 if and only if $\mathbf{b}$ is contained in $\text{span}(\mathbf{a}_1, \cdots, \mathbf{a}_n)$. In words, there is a solution to 7.5 if and only if $\mathbf{b}$ is in the column space of A. In terms of rank, the following proposition describes the situation.

Proposition 7.4.26 *Let A be an $m \times n$ matrix and let $\mathbf{b}$ be an $m \times 1$ column vector. Then there exists a solution to 7.5 if and only if*

$$\text{rank}\begin{pmatrix} A & | & \mathbf{b} \end{pmatrix} = \text{rank}(A). \tag{7.6}$$

Proof: Place $\begin{pmatrix} A & | & \mathbf{b} \end{pmatrix}$ and A in row reduced echelon form, respectively B and C. If the above condition on rank is true, then both B and C have the same number of nonzero rows. In particular, you cannot have a row of the form

$$\begin{pmatrix} 0 & \cdots & 0 & \blacksquare \end{pmatrix}$$

where $\blacksquare \neq 0$ in B. Therefore, there will exist a solution to the system 7.5.

Conversely, suppose there exists a solution. This means there cannot be such a row in B described above. Therefore, B and C must have the same number of zero rows and so they have the same number of nonzero rows. Therefore, the rank of the two matrices in 7.6 is the same. This proves the proposition.

7.5 Fredholm Alternative

There is a very useful version of Proposition 7.4.26 known as the **Fredholm alternative**. I will only present this for the case of real matrices here. Later a much more elegant and general approach is presented which allows for the general case of complex matrices.

The following definition is used to state the Fredholm alternative.

Definition 7.5.1 *Let $S \subseteq \mathbb{R}^m$. Then $S^\perp \equiv \{\mathbf{z} \in \mathbb{R}^m : \mathbf{z} \cdot \mathbf{s} = 0$ for every $\mathbf{s} \in S\}$. The funny exponent, $\perp$ is called "perp".*

[2] Isn't it amazing how many different words are available for use in linear algebra?

Now note

$$\ker\left(A^T\right) \equiv \left\{\mathbf{z} : A^T\mathbf{z} = \mathbf{0}\right\} = \left\{\mathbf{z} : \sum_{k=1}^{m} z_k \mathbf{a}_k = 0\right\}$$

Lemma 7.5.2 *Let A be a real $m \times n$ matrix, let $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$. Then*

$$\left(A\mathbf{x} \cdot \mathbf{y}\right) = \left(\mathbf{x} \cdot A^T\mathbf{y}\right)$$

Proof: This follows right away from the definition of the dot product and matrix multiplication.

$$\begin{aligned}\left(A\mathbf{x} \cdot \mathbf{y}\right) &= \sum_{k,l} A_{kl} x_l y_k \\ &= \sum_{k,l} \left(A^T\right)_{lk} x_l y_k \\ &= \left(\mathbf{x} \cdot A^T\mathbf{y}\right).\end{aligned}$$

This proves the lemma.

Now it is time to state the Fredholm alternative. The first version of this is the following theorem.

Theorem 7.5.3 *Let A be a real $m \times n$ matrix and let $\mathbf{b} \in \mathbb{R}^m$. There exists a solution, $\mathbf{x}$ to the equation $A\mathbf{x} = \mathbf{b}$ if and only if $\mathbf{b} \in \ker\left(A^T\right)^{\perp}$.*

Proof: First suppose $\mathbf{b} \in \ker\left(A^T\right)^{\perp}$. Then this says that if $A^T\mathbf{x} = \mathbf{0}$, it follows that $\mathbf{b} \cdot \mathbf{x} = \mathbf{0}$. In other words, taking the transpose, if

$$\mathbf{x}^T A = \mathbf{0}, \text{ then } \mathbf{b} \cdot \mathbf{x} = 0.$$

In other words, letting $\mathbf{x} = (x_1, \cdots, x_m)^T$, it follows that if

$$\sum_{i=1}^{m} x_i A_{ij} = 0 \text{ for each } j,$$

then it follows

$$\sum_i b_i x_i = 0.$$

In other words, if you get a row of zeros in row reduced echelon form for A then you the same row operations produce a zero in the $m \times 1$ matrix $\mathbf{b}$.

Consequently

$$\operatorname{rank}\left(\begin{array}{c|c} A & \mathbf{b} \end{array}\right) = \operatorname{rank}(A)$$

and so by Proposition 7.4.26, there exists a solution, $\mathbf{x}$ to the system $A\mathbf{x} = \mathbf{b}$. It remains to go the other direction.

Let $\mathbf{z} \in \ker\left(A^T\right)$ and suppose $A\mathbf{x} = \mathbf{b}$. I need to verify $\mathbf{b} \cdot \mathbf{z} = 0$. By Lemma 7.5.2,

$$\mathbf{b} \cdot \mathbf{z} = A\mathbf{x} \cdot \mathbf{z} = \mathbf{x} \cdot A^T\mathbf{z} = \mathbf{x} \cdot \mathbf{0} = 0$$

This proves the theorem.

This implies the following corollary which is also called the Fredholm alternative. The "alternative" becomes more clear in this corollary.

Corollary 7.5.4 *Let A be an $m \times n$ matrix. Then A maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ if and only if the only solution to $A^T\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$.*

Proof: If the only solution to $A^T\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$, then $\ker\left(A^T\right) = \{\mathbf{0}\}$ and so $\ker\left(A^T\right)^\perp = \mathbb{R}^m$ because every $\mathbf{b} \in \mathbb{R}^m$ has the property that $\mathbf{b} \cdot \mathbf{0} = 0$. Therefore, $A\mathbf{x} = \mathbf{b}$ has a solution for any $\mathbf{b} \in \mathbb{R}^m$ because the $\mathbf{b}$ for which there is a solution are those in $\ker\left(A^T\right)^\perp$ by Theorem 7.5.3. In other words, A maps $\mathbb{R}^n$ onto $\mathbb{R}^m$.

Conversely if A is onto, then by Theorem 7.5.3 every $\mathbf{b} \in \mathbb{R}^m$ is in $\ker\left(A^T\right)^\perp$ and so if $A^T\mathbf{x} = \mathbf{0}$, then $\mathbf{b} \cdot \mathbf{x} = 0$ for every $\mathbf{b}$. In particular, this holds for $\mathbf{b} = \mathbf{x}$. Hence if $A^T\mathbf{x} = \mathbf{0}$, then $\mathbf{x} = \mathbf{0}$. This proves the corollary.

Here is an amusing example.

Example 7.5.5 *Let A be an $m \times n$ matrix in which $m > n$. Then A cannot map onto $\mathbb{R}^m$.*

The reason for this is that A^T is an $n \times m$ where $m > n$ and so in the augmented matrix,

$$\left(A^T|\mathbf{0}\right)$$

there must be some free variables. Thus there exists a nonzero vector $\mathbf{x}$ such that $A^T\mathbf{x} = \mathbf{0}$.

7.5.1 Row, Column, And Determinant Rank

I will now present a review of earlier topics and prove Theorem 7.3.4.

Definition 7.5.6 *A sub-matrix of a matrix A is the rectangular array of numbers obtained by deleting some rows and columns of A. Let A be an $m \times n$ matrix. The **determinant rank** of the matrix equals r where r is the largest number such that some $r \times r$ sub-matrix of A has a non zero determinant. The **row rank** is defined to be the dimension of the span of the rows. The **column rank** is defined to be the dimension of the span of the columns.*

Theorem 7.5.7 *If A, an $m \times n$ matrix has determinant rank, r, then there exist r rows of the matrix such that every other row is a linear combination of these r rows.*

Proof: Suppose the determinant rank of $A = (a_{ij})$ equals r. Thus some $r \times r$ submatrix has non zero determinant and there is no larger square submatrix which has non zero determinant. Suppose such a submatrix is determined by the r columns whose indices are

$$j_1 < \cdots < j_r$$

and the r rows whose indices are

$$i_1 < \cdots < i_r$$

I want to show that every row is a linear combination of these rows. Consider the l^{th} row and let p be an index between 1 and n. Form the following $(r+1) \times (r+1)$ matrix

$$\begin{pmatrix} a_{i_1 j_1} & \cdots & a_{i_1 j_r} & a_{i_1 p} \\ \vdots & & \vdots & \vdots \\ a_{i_r j_1} & \cdots & a_{i_r j_r} & a_{i_r p} \\ a_{l j_1} & \cdots & a_{l j_r} & a_{l p} \end{pmatrix}$$

Of course you can assume $l \notin \{i_1, \cdots, i_r\}$ because there is nothing to prove if the l^{th} row is one of the chosen ones. The above matrix has determinant 0. This is because if $p \notin \{j_1, \cdots, j_r\}$ then the above would be a submatrix of A which is too large to have non

zero determinant. On the other hand, if $p \in \{j_1, \cdots, j_r\}$ then the above matrix has two columns which are equal so its determinant is still 0.

Expand the determinant of the above matrix along the last column. Let C_k denote the cofactor associated with the entry $a_{i_k p}$. This is not dependent on the choice of p. Remember, you delete the column and the row the entry is in and take the determinant of what is left and multiply by -1 raised to an appropriate power. Let C denote the cofactor associated with a_{lp}. This is given to be nonzero, it being the determinant of the matrix

$$\begin{pmatrix} a_{i_1 j_1} & \cdots & a_{i_1 j_r} \\ \vdots & & \vdots \\ a_{i_r j_1} & \cdots & a_{i_r j_r} \end{pmatrix}$$

Thus

$$0 = a_{lp} C + \sum_{k=1}^{r} C_k a_{i_k p}$$

which implies

$$a_{lp} = \sum_{k=1}^{r} \frac{-C_k}{C} a_{i_k p} \equiv \sum_{k=1}^{r} m_k a_{i_k p}$$

Since this is true for every p and since m_k does not depend on p, this has shown the l^{th} row is a linear combination of the $i_1, i_2, \cdots, i_r$ rows. This proves the theorem.

Corollary 7.5.8 *The determinant rank equals the row rank.*

Proof: From Theorem 7.5.7, the row rank is no larger than the determinant rank. Could the row rank be smaller than the determinant rank? If so, there exist p rows for $p < r$ such that the span of these p rows equals the row space. But this implies that the $r \times r$ sub-matrix whose determinant is nonzero also has row rank no larger than p which is impossible if its determinant is to be nonzero because at least one row is a linear combination of the others.

Corollary 7.5.9 *If A has determinant rank, r, then there exist r columns of the matrix such that every other column is a linear combination of these r columns. Also the column rank equals the determinant rank.*

Proof: This follows from the above by considering A^T. The rows of A^T are the columns of A and the determinant rank of A^T and A are the same. Therefore, from Corollary 7.5.8, column rank of A = row rank of A^T = determinant rank of A^T = determinant rank of A.

The following theorem is of fundamental importance and ties together many of the ideas presented above.

Theorem 7.5.10 *Let A be an $n \times n$ matrix. Then the following are equivalent.*

1. $\det(A) = 0$.
2. *A, A^T are not one to one.*
3. *A is not onto.*

Proof: Suppose $\det(A) = 0$. Then the determinant rank of $A = r < n$. Therefore, there exist r columns such that every other column is a linear combination of these columns by Theorem 7.5.7. In particular, it follows that for some m, the m^{th} column is a linear

combination of all the others. Thus letting $A = \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_m & \cdots & \mathbf{a}_n \end{pmatrix}$ where the columns are denoted by $\mathbf{a}_i$, there exists scalars, α_i such that

$$\mathbf{a}_m = \sum_{k \neq m} \alpha_k \mathbf{a}_k.$$

Now consider the column vector, $\mathbf{x} \equiv \begin{pmatrix} \alpha_1 & \cdots & -1 & \cdots & \boldsymbol{\alpha}_n \end{pmatrix}^T$. Then

$$A\mathbf{x} = -\mathbf{a}_m + \sum_{k \neq m} \alpha_k \mathbf{a}_k = \mathbf{0}.$$

Since also $A\mathbf{0} = \mathbf{0}$, it follows A is not one to one. Similarly, A^T is not one to one by the same argument applied to A^T. This verifies that 1.) implies 2.).

Now suppose 2.). Then since A^T is not one to one, it follows there exists $\mathbf{x} \neq \mathbf{0}$ such that

$$A^T\mathbf{x} = \mathbf{0}.$$

Taking the transpose of both sides yields

$$\mathbf{x}^T A = \mathbf{0}$$

where the $\mathbf{0}$ is a $1 \times n$ matrix or row vector. Now if $A\mathbf{y} = \mathbf{x}$, then

$$|\mathbf{x}|^2 = \mathbf{x}^T (A\mathbf{y}) = \left(\mathbf{x}^T A\right) \mathbf{y} = \mathbf{0}\mathbf{y} = 0$$

contrary to $\mathbf{x} \neq \mathbf{0}$. Consequently there can be no $\mathbf{y}$ such that $A\mathbf{y} = \mathbf{x}$ and so A is not onto. This shows that 2.) implies 3.).

Finally, suppose 3.). If 1.) does not hold, then $\det(A) \neq 0$ but then from Theorem 6.4.15 A^{-1} exists and so for every $\mathbf{y} \in \mathbb{F}^n$ there exists a unique $\mathbf{x} \in \mathbb{F}^n$ such that $A\mathbf{x} = \mathbf{y}$. In fact $\mathbf{x} = A^{-1}\mathbf{y}$. Thus A would be onto contrary to 3.). This shows 3.) implies 1.) and proves the theorem.

Corollary 7.5.11 *Let A be an $n \times n$ matrix. Then the following are equivalent.*

1. *det$(A) \neq 0$.*
2. *A and A^T are one to one.*
3. *A is onto.*

Proof: This follows immediately from the above theorem.

Corollary 7.5.12 *Let A be an invertible $n \times n$ matrix. Then A equals a finite product of elementary matrices.*

Proof: Since A^{-1} is given to exist, $\det(A) \neq 0$ and it follows A must have rank n and so the row reduced echelon form of A is I. Therefore, by Theorem 7.1.6 there is a sequence of elementary matrices, $E_1, \cdots, E_p$ which accomplish successive row operations such that

$$(E_p E_{p-1} \cdots E_1) A = I.$$

But now multiply on the left on both sides by E_p^{-1} then by E_{p-1}^{-1} and then by E_{p-2}^{-1} etc. until you get

$$A = E_1^{-1} E_2^{-1} \cdots E_{p-1}^{-1} E_p^{-1}$$

and by Theorem 7.1.6 each of these in this product is an elementary matrix.

7.6 Linear Transformations

An $m \times n$ matrix can be used to transform vectors in $\mathbb{F}^n$ to vectors in $\mathbb{F}^m$ through the use of matrix multiplication.

Example 7.6.1 *Consider the matrix,* $\begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{pmatrix}$. *Think of it as a function which takes vectors in* $\mathbb{F}^3$ *and makes them in to vectors in* $\mathbb{F}^2$ *as follows. For* $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ *a vector in* $\mathbb{F}^3$, *multiply on the left by the given matrix to obtain the vector in* $\mathbb{F}^2$. *Here are some numerical examples.*

$$\begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix} = \begin{pmatrix} -3 \\ 0 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 10 \\ 5 \\ -3 \end{pmatrix} = \begin{pmatrix} 20 \\ 25 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 7 \\ 3 \end{pmatrix} = \begin{pmatrix} 14 \\ 7 \end{pmatrix},$$

More generally,

$$\begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x + 2y \\ 2x + y \end{pmatrix}$$

The idea is to define a function which takes vectors in $\mathbb{F}^3$ *and delivers new vectors in* $\mathbb{F}^2$.

This is an example of something called a linear transformation.

Definition 7.6.2 *Let* $T : \mathbb{F}^n \to \mathbb{F}^m$ *be a function. Thus for each* $\mathbf{x} \in \mathbb{F}^n, T\mathbf{x} \in \mathbb{F}^m$. *Then* T *is a* ***linear transformation*** *if whenever* α, β *are scalars and* $\mathbf{x}_1$ *and* $\mathbf{x}_2$ *are vectors in* $\mathbb{F}^n$,

$$T(\alpha\mathbf{x}_1 + \beta\mathbf{x}_2) = \alpha_1 T\mathbf{x}_1 + \beta T\mathbf{x}_2.$$

In words, linear transformations distribute across + and allow you to factor out scalars. At this point, recall the properties of matrix multiplication. The pertinent property is 4.14 on Page 49. Recall it states that for a and b scalars,

$$A(aB + bC) = aAB + bAC$$

In particular, for A an $m \times n$ matrix and B and $C, n \times 1$ matrices (column vectors) the above formula holds which is nothing more than the statement that matrix multiplication gives an example of a linear transformation.

Definition 7.6.3 *A linear transformation is called* ***one to one*** *(often written as* $1-1$*) if it never takes two different vectors to the same vector. Thus* T *is one to one if whenever* $\mathbf{x} \neq \mathbf{y}$

$$T\mathbf{x} \neq T\mathbf{y}.$$

Equivalently, if $T(\mathbf{x}) = T(\mathbf{y})$, *then* $\mathbf{x} = \mathbf{y}$.

In the case that a linear transformation comes from matrix multiplication, it is common usage to refer to the matrix as a one to one matrix when the linear transformation it determines is one to one.

Definition 7.6.4 *A linear transformation mapping $\mathbb{F}^n$ to $\mathbb{F}^m$ is called* ***onto*** *if whenever $\mathbf{y} \in \mathbb{F}^m$ there exists $\mathbf{x} \in \mathbb{F}^n$ such that $T(\mathbf{x}) = \mathbf{y}$.*

Thus T is onto if everything in $\mathbb{F}^m$ gets hit. In the case that a linear transformation comes from matrix multiplication, it is common to refer to the matrix as onto when the linear transformation it determines is onto. Also it is common usage to write $T\mathbb{F}^n$, $T(\mathbb{F}^n)$,or $\mathrm{Im}(T)$ as the set of vectors of $\mathbb{F}^m$ which are of the form $T\mathbf{x}$ for some $\mathbf{x} \in \mathbb{F}^n$. In the case that T is obtained from multiplication by an $m \times n$ matrix, A, it is standard to simply write $A(\mathbb{F}^n)$ $A\mathbb{F}^n$, or $\mathrm{Im}(A)$ to denote those vectors in $\mathbb{F}^m$ which are obtained in the form $A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{F}^n$.

7.7 Constructing The Matrix Of A Linear Transformation

It turns out that if T is any linear transformation which maps $\mathbb{F}^n$ to $\mathbb{F}^m$, there is always an $m \times n$ matrix, A with the property that

$$A\mathbf{x} = T\mathbf{x} \tag{7.7}$$

for all $\mathbf{x} \in \mathbb{F}^n$. Here is why. Suppose $T : \mathbb{F}^n \to \mathbb{F}^m$ is a linear transformation and you want to find the matrix defined by this linear transformation as described in 7.7. Then if $\mathbf{x} \in \mathbb{F}^n$ it follows

$$\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{e}_i$$

where $\mathbf{e}_i$ is the vector which has zeros in every slot but the i^{th} and a 1 in this slot. Then since T is linear,

$$\begin{aligned} T\mathbf{x} &= \sum_{i=1}^{n} x_i T(\mathbf{e}_i) \\ &= \begin{pmatrix} | & & | \\ T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \\ | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\ &\equiv A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \end{aligned}$$

and so you see that the matrix desired is obtained from letting the i^{th} column equal $T(\mathbf{e}_i)$. We state this as the following theorem.

Theorem 7.7.1 *Let T be a linear transformation from $\mathbb{F}^n$ to $\mathbb{F}^m$. Then the matrix, A satisfying 7.7 is given by*

$$\begin{pmatrix} | & & | \\ T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \\ | & & | \end{pmatrix}$$

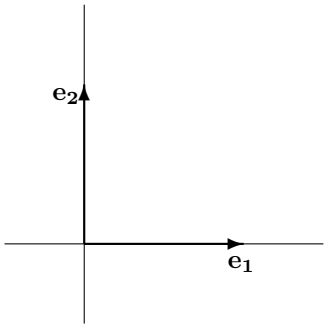
where $T\mathbf{e}_i$ is the i^{th} column of A.

7.7.1 Rotations of $\mathbb{R}^2$

Sometimes you need to find a matrix which represents a given linear transformation which is described in geometrical terms. The idea is to produce a matrix which you can multiply a vector by to get the same thing as some geometrical description. A good example of this is the problem of rotation of vectors.

Example 7.7.2 *Determine the matrix which represents the linear transformation defined by rotating every vector through an angle of θ.*

Let $\mathbf{e}_1 \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\mathbf{e}_2 \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. These identify the geometric vectors which point along the positive x axis and positive y axis as shown.



From the above, you only need to find $T\mathbf{e}_1$ and $T\mathbf{e}_2$, the first being the first column of the desired matrix, A and the second being the second column. From drawing a picture and doing a little geometry, you see that

$$T\mathbf{e}_1 = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, T\mathbf{e}_2 = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}.$$

Therefore, from Theorem 7.7.1,

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Example 7.7.3 *Find the matrix of the linear transformation which is obtained by first rotating all vectors through an angle of ϕ and then through an angle θ. Thus you want the linear transformation which rotates all angles through an angle of $\theta + \phi$.*

Let $T_{\theta+\phi}$ denote the linear transformation which rotates every vector through an angle of $\theta + \phi$. Then to get $T_{\theta+\phi}$, you could first do T_ϕ and then do T_θ where T_ϕ is the linear transformation which rotates through an angle of ϕ and T_θ is the linear transformation which rotates through an angle of θ. Denoting the corresponding matrices by $A_{\theta+\phi}$, A_ϕ, and A_θ, you must have for every $\mathbf{x}$

$$A_{\theta+\phi}\mathbf{x} = T_{\theta+\phi}\mathbf{x} = T_\theta T_\phi \mathbf{x} = A_\theta A_\phi \mathbf{x}.$$

Consequently, you must have

$$\begin{aligned} A_{\theta+\phi} &= \begin{pmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{pmatrix} = A_\theta A_\phi \\ &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}. \end{aligned}$$

You know how to multiply matrices. Do so to the pair on the right. This yields

$$\begin{pmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{pmatrix} = \begin{pmatrix} \cos\theta\cos\phi - \sin\theta\sin\phi & -\cos\theta\sin\phi - \sin\theta\cos\phi \\ \sin\theta\cos\phi + \cos\theta\sin\phi & \cos\theta\cos\phi - \sin\theta\sin\phi \end{pmatrix}.$$

Don't these look familiar? They are the usual trig. identities for the sum of two angles derived here using linear algebra concepts.

You do not have to stop with two dimensions. You can consider rotations and other geometric concepts in any number of dimensions. This is one of the major advantages of linear algebra. You can break down a difficult geometrical procedure into small steps, each corresponding to multiplication by an appropriate matrix. Then by multiplying the matrices, you can obtain a single matrix which can give you numerical information on the results of applying the given sequence of simple procedures. That which you could never visualize can still be understood to the extent of finding exact numerical answers. Another example follows.

Example 7.7.4 *Find the matrix of the linear transformation which is obtained by first rotating all vectors through an angle of $\pi/6$ and then reflecting through the x axis.*

As shown in Example 7.7.3, the matrix of the transformation which involves rotating through an angle of $\pi/6$ is

$$\begin{pmatrix} \cos(\pi/6) & -\sin(\pi/6) \\ \sin(\pi/6) & \cos(\pi/6) \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\sqrt{3} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2}\sqrt{3} \end{pmatrix}$$

The matrix for the transformation which reflects all vectors through the x axis is

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Therefore, the matrix of the linear transformation which first rotates through $\pi/6$ and then reflects through the x axis is

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} \frac{1}{2}\sqrt{3} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2}\sqrt{3} \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\sqrt{3} & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2}\sqrt{3} \end{pmatrix}.$$

7.7.2 Projections

In Physics it is important to consider the work done by a force field on an object. This involves the concept of projection onto a vector. Suppose you want to find the projection of a vector, $\mathbf{v}$ onto the given vector, $\mathbf{u}$, denoted by $\mathrm{proj}_{\mathbf{u}}(\mathbf{v})$ This is done using the dot product as follows.

$$\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) = \left(\frac{\mathbf{v}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u}$$

Because of properties of the dot product, the map $\mathbf{v} \to \mathrm{proj}_{\mathbf{u}}(\mathbf{v})$ is linear,

$$\begin{aligned} \mathrm{proj}_{\mathbf{u}}(\alpha\mathbf{v}+\beta\mathbf{w}) &= \left(\frac{\alpha\mathbf{v}+\beta\mathbf{w}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u} = \alpha\left(\frac{\mathbf{v}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u} + \beta\left(\frac{\mathbf{w}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u} \\ &= \alpha\,\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) + \beta\,\mathrm{proj}_{\mathbf{u}}(\mathbf{w}). \end{aligned}$$

Example 7.7.5 *Let the projection map be defined above and let $\mathbf{u} = (1,2,3)^T$. Does this linear transformation come from multiplication by a matrix? If so, what is the matrix?*

You can find this matrix in the same way as in the previous example. Let $\mathbf{e}_i$ denote the vector in $\mathbb{R}^n$ which has a 1 in the i^{th} position and a zero everywhere else. Thus a typical vector, $\mathbf{x} = (x_1, \cdots, x_n)^T$ can be written in a unique way as

$$\mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{e}_j.$$

From the way you multiply a matrix by a vector, it follows that $\mathrm{proj}_{\mathbf{u}}(\mathbf{e}_i)$ gives the i^{th} column of the desired matrix. Therefore, it is only necessary to find

$$\mathrm{proj}_{\mathbf{u}}(\mathbf{e}_i) \equiv \left(\frac{\mathbf{e}_i \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\right) \mathbf{u}$$

For the given vector in the example, this implies the columns of the desired matrix are

$$\frac{1}{14}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \frac{2}{14}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \frac{3}{14}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Hence the matrix is

$$\frac{1}{14}\begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix}.$$

7.7.3 Matrices Which Are One To One Or Onto

Lemma 7.7.6 *Let A be an $m \times n$ matrix. Then $A(\mathbb{F}^n) = \mathrm{span}(\mathbf{a}_1, \cdots, \mathbf{a}_n)$ where $\mathbf{a}_1, \cdots, \mathbf{a}_n$ denote the columns of A. In fact, for $\mathbf{x} = (x_1, \cdots, x_n)^T$,*

$$A\mathbf{x} = \sum_{k=1}^{n} x_k \mathbf{a}_k.$$

Proof: This follows from the definition of matrix multiplication in Definition 4.1.9 on Page 44.

The following is a theorem of major significance. First here is an interesting observation.

Observation 7.7.7 *Let A be an $m \times n$ matrix. Then A is one to one if and only if $A\mathbf{x} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$.*

Here is why: $A\mathbf{0} = A(\mathbf{0} + \mathbf{0}) = A\mathbf{0} + A\mathbf{0}$ and so $A\mathbf{0} = \mathbf{0}$.

Now suppose A is one to one and $A\mathbf{x} = \mathbf{0}$. Then since $A\mathbf{0} = \mathbf{0}$, it follows $\mathbf{x} = \mathbf{0}$. Thus if A is one to one and $A\mathbf{x} = \mathbf{0}$, then $\mathbf{x} = \mathbf{0}$.

Next suppose the condition that $A\mathbf{x} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$ is valid. Then if $A\mathbf{x} = A\mathbf{y}$, then $A(\mathbf{x} - \mathbf{y}) = \mathbf{0}$ and so from the condition, $\mathbf{x} - \mathbf{y} = \mathbf{0}$ so that $\mathbf{x} = \mathbf{y}$. Thus A is one to one.

Theorem 7.7.8 *Suppose A is an $n \times n$ matrix. Then A is one to one if and only if A is onto. Also, if B is an $n \times n$ matrix and $AB = I$, then it follows $BA = I$.*

Proof: First suppose A is one to one. Consider the vectors, $\{A\mathbf{e}_1, \cdots, A\mathbf{e}_n\}$ where $\mathbf{e}_k$ is the column vector which is all zeros except for a 1 in the k^{th} position. This set of vectors is linearly independent because if

$$\sum_{k=1}^{n} c_k A\mathbf{e}_k = \mathbf{0},$$

then since A is linear,

$$A\left(\sum_{k=1}^{n} c_k \mathbf{e}_k\right) = \mathbf{0}$$

and since A is one to one, it follows

$$\sum_{k=1}^{n} c_k \mathbf{e}_k = \mathbf{0}$$

which implies each $c_k = 0$. Therefore, $\{A\mathbf{e}_1, \cdots, A\mathbf{e}_n\}$ must be a basis for $\mathbb{F}^n$ by Corollary 7.4.16 on Page 124. It follows that for $\mathbf{y} \in \mathbb{F}^n$ there exist constants, c_i such that

$$\mathbf{y} = \sum_{k=1}^{n} c_k A\mathbf{e}_k = A\left(\sum_{k=1}^{n} c_k \mathbf{e}_k\right)$$

showing that, since $\mathbf{y}$ was arbitrary, A is onto.

Next suppose A is onto. This implies the span of the columns of A equals $\mathbb{F}^n$ and by Corollary 7.4.16 this implies the columns of A are independent. If $A\mathbf{x} = \mathbf{0}$, then letting $\mathbf{x} = (x_1, \cdots, x_n)^T$, it follows

$$\sum_{i=1}^{n} x_i \mathbf{a}_i = \mathbf{0}$$

and so each $x_i = 0$. If $A\mathbf{x} = A\mathbf{y}$, then $A(\mathbf{x} - \mathbf{y}) = \mathbf{0}$ and so $\mathbf{x} = \mathbf{y}$. This shows A is one to one.

Now suppose $AB = I$. Why is $BA = I$? Since $AB = I$ it follows B is one to one since otherwise, there would exist, $\mathbf{x} \neq \mathbf{0}$ such that $B\mathbf{x} = \mathbf{0}$ and then $AB\mathbf{x} = A\mathbf{0} = \mathbf{0} \neq I\mathbf{x}$. Therefore, from what was just shown, B is also onto. In addition to this, A must be one to one because if $A\mathbf{y} = \mathbf{0}$, then $\mathbf{y} = B\mathbf{x}$ for some $\mathbf{x}$ and then $\mathbf{x} = AB\mathbf{x} = A\mathbf{y} = \mathbf{0}$ showing $\mathbf{y} = \mathbf{0}$. Now from what is given to be so, it follows $(AB)A = A$ and so using the associative law for matrix multiplication,

$$A(BA) - A = A(BA - I) = 0.$$

But this means $(BA - I)\mathbf{x} = \mathbf{0}$ for all $\mathbf{x}$ since otherwise, A would not be one to one. Hence $BA = I$ as claimed. This proves the theorem.

This theorem shows that if an $n \times n$ matrix, B acts like an inverse when multiplied on one side of A it follows that $B = A^{-1}$and it will act like an inverse on both sides of A.

The conclusion of this theorem pertains to square matrices only. For example, let

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}, \; B = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & -1 \end{pmatrix} \tag{7.8}$$

Then

$$BA = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

but

$$AB = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & -1 \\ 1 & 0 & 0 \end{pmatrix}.$$

There is also an important characterization in terms of determinants. This is proved completely in the section on the mathematical theory of the determinant.

Theorem 7.7.9 *Let A be an $n \times n$ matrix and let T_A denote the linear transformation determined by A. Then the following are equivalent.*

1. *T_A is one to one.*
2. *T_A is onto.*
3. $\det(A) \neq 0$.

7.7.4 The General Solution Of A Linear System

Recall the following definition which was discussed above.

Definition 7.7.10 *T is a **linear transformation** if whenever $\mathbf{x}, \mathbf{y}$ are vectors and a, b scalars,*

$$T(a\mathbf{x} + b\mathbf{y}) = aT\mathbf{x} + bT\mathbf{y}. \tag{7.9}$$

Thus linear transformations distribute across addition and pass scalars to the outside. A linear system is one which is of the form

$$T\mathbf{x} = \mathbf{b}.$$

*If $T\mathbf{x}_p = \mathbf{b}$, then $\mathbf{x}_p$ is called a **particular solution** to the linear system.*

For example, if A is an $m \times n$ matrix and T_A is determined by

$$T_A(\mathbf{x}) = A\mathbf{x},$$

then from the properties of matrix multiplication, T_A is a linear transformation. In this setting, we will usually write A for the linear transformation as well as the matrix. There are many other examples of linear transformations other than this. In differential equations, you will encounter linear transformations which act on functions to give new functions. In this case, the functions are considered as vectors. Don't worry too much about this at this time. It will happen later. The fundamental idea is that something is linear if 7.9 holds and if whenever a, b are scalars and $\mathbf{x}, \mathbf{y}$ are vectors $a\mathbf{x} + b\mathbf{y}$ is a vector. That is you can add vectors and multiply by scalars.

Definition 7.7.11 *Let T be a linear transformation. Define*

$$\ker(T) \equiv \{\mathbf{x} : T\mathbf{x} = \mathbf{0}\}.$$

*In words, $\ker(T)$ is called the **kernel** of T. As just described, $\ker(T)$ consists of the set of all vectors which T sends to $\mathbf{0}$. This is also called the **null space** of T. It is also called the **solution space** of the equation $T\mathbf{x} = \mathbf{0}$.*

The above definition states that $\ker(T)$ is the set of solutions to the equation,

$$T\mathbf{x} = \mathbf{0}.$$

In the case where T is really a matrix, you have been solving such equations for quite some time. However, sometimes linear transformations act on vectors which are not in $\mathbb{F}^n$. There is more on this in Chapter 14 on Page 14 and this is discussed more carefully then. However, consider the following familiar example.

Example 7.7.12 *Let $\frac{d}{dx}$ denote the linear transformation defined on X, the functions which are defined on $\mathbb{R}$ and have a continuous derivative. Find $\ker\left(\frac{d}{dx}\right)$.*

The example asks for functions, f which the property that $\frac{df}{dx} = 0$. As you know from calculus, these functions are the constant functions. Thus $\ker\left(\frac{d}{dx}\right)$ = constant functions.

When T is a linear transformation, systems of the form $T\mathbf{x} = \mathbf{0}$ are called **homogeneous systems**. Thus the solution to the homogeneous system is known as $\ker(T)$.

Systems of the form $T\mathbf{x} = \mathbf{b}$ where $\mathbf{b} \neq \mathbf{0}$ are called **nonhomogeneous systems**. It turns out there is a very interesting and important relation between the solutions to the homogeneous systems and the solutions to the nonhomogeneous systems.

Theorem 7.7.13 *Suppose $\mathbf{x}_p$ is a solution to the linear system,*

$$T\mathbf{x} = \mathbf{b}$$

Then if $\mathbf{y}$ is any other solution to the linear system, there exists $\mathbf{x} \in \ker(T)$ such that

$$\mathbf{y} = \mathbf{x}_p + \mathbf{x}.$$

Proof: Consider $\mathbf{y} - \mathbf{x}_p \equiv \mathbf{y} + (-1)\mathbf{x}_p$. Then $T\left(\mathbf{y} - \mathbf{x}_p\right) = T\mathbf{y} - T\mathbf{x}_p = \mathbf{b} - \mathbf{b} = \mathbf{0}$. Let $\mathbf{x} \equiv \mathbf{y} - \mathbf{x}_p$. This proves the theorem.

Sometimes people remember the above theorem in the following form. The solutions to the nonhomogeneous system, $T\mathbf{x} = \mathbf{b}$ are given by $\mathbf{x}_p + \ker(T)$ where $\mathbf{x}_p$ is a particular solution to $T\mathbf{x} = \mathbf{b}$.

We have been vague about what T is and what $\mathbf{x}$ is on purpose. This theorem is completely algebraic in nature and will work whenever you have linear transformations. In particular, it will be important in differential equations. For now, here is a familiar example.

Example 7.7.14 *Let*

$$A = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 2 & 1 & 1 & 2 \\ 4 & 5 & 7 & 2 \end{pmatrix}$$

Find $\ker(A)$. *Equivalently, find the solution space to the system of equations $A\mathbf{x} = \mathbf{0}$.*

This asks you to find $\{\mathbf{x} : A\mathbf{x} = \mathbf{0}\}$. In other words you are asked to solve the system, $A\mathbf{x} = \mathbf{0}$. Let $\mathbf{x} = (x, y, z, w)^T$. Then this amounts to solving

$$\begin{pmatrix} 1 & 2 & 3 & 0 \\ 2 & 1 & 1 & 2 \\ 4 & 5 & 7 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

This is the linear system

$$\begin{gathered} x + 2y + 3z = 0 \\ 2x + y + z + 2w = 0 \\ 4x + 5y + 7z + 2w = 0 \end{gathered}$$

and you know how to solve this using row operations, (Gauss Elimination). Set up the augmented matrix,

$$\left(\begin{array}{cccc|c} 1 & 2 & 3 & 0 & 0 \\ 2 & 1 & 1 & 2 & 0 \\ 4 & 5 & 7 & 2 & 0 \end{array}\right)$$

Then row reduce to obtain the row reduced echelon form,

$$\left(\begin{array}{cccc|c} 1 & 0 & -\frac{1}{3} & \frac{4}{3} & 0 \\ 0 & 1 & \frac{5}{3} & -\frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

This yields $x = \frac{1}{3}z - \frac{4}{3}w$ and $y = \frac{2}{3}w - \frac{5}{3}z$. Thus $\ker(A)$ consists of vectors of the form,

$$\begin{pmatrix} \frac{1}{3}z - \frac{4}{3}w \\ \frac{2}{3}w - \frac{5}{3}z \\ z \\ w \end{pmatrix} = z \begin{pmatrix} \frac{1}{3} \\ -\frac{5}{3} \\ 1 \\ 0 \end{pmatrix} + w \begin{pmatrix} -\frac{4}{3} \\ \frac{2}{3} \\ 0 \\ 1 \end{pmatrix}.$$

Example 7.7.15 *The **general solution** of a linear system of equations is just the set of all solutions. Find the general solution to the linear system,*

$$\begin{pmatrix} 1 & 2 & 3 & 0 \\ 2 & 1 & 1 & 2 \\ 4 & 5 & 7 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 9 \\ 7 \\ 25 \end{pmatrix}$$

given that $\begin{pmatrix} 1 & 1 & 2 & 1 \end{pmatrix}^T = \begin{pmatrix} x & y & z & w \end{pmatrix}^T$ *is one solution.*

Note the matrix on the left is the same as the matrix in Example 7.7.14. Therefore, from Theorem 7.7.13, you will obtain all solutions to the above linear system in the form

$$z \begin{pmatrix} \frac{1}{3} \\ -\frac{5}{3} \\ 1 \\ 0 \end{pmatrix} + w \begin{pmatrix} -\frac{4}{3} \\ \frac{2}{3} \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 2 \\ 1 \end{pmatrix}$$

because $\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 2 \\ 1 \end{pmatrix}$ is a particular solution to the given system of equations.

The LU Factorization

8.0.5 Outcomes

A. Determine LU factorizations when possible.

B. Solve a linear system of equations using the LU factorization.

8.1 Definition Of An LU factorization

An LU factorization of a matrix involves writing the given matrix as the product of a lower triangular matrix which has the main diagonal consisting entirely of ones L, and an upper triangular matrix U in the indicated order. This is the version discussed here but it is sometimes the case that the L has numbers other than 1 down the main diagonal. It is still a useful concept. The L goes with "lower" and the U with "upper". It turns out many matrices can be written in this way and when this is possible, people get excited about slick ways of solving the system of equations, $A\mathbf{x} = \mathbf{y}$. It is for this reason that you want to study the LU factorization. It allows you to work only with triangular matrices. It turns out that it takes about $2n^3/3$ operations to use Gauss elimination but only $n^3/3$ to obtain an LU factorization.

First it should be noted not all matrices have an LU factorization and so we will emphasize the techniques for achieving it rather than formal proofs.

Example 8.1.1 *Can you write* $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ *in the form* LU *as just described?*

To do so you would need

$$\begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix} \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} = \begin{pmatrix} a & b \\ xa & xb+c \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Therefore, $b = 1$ and $a = 0$. Also, from the bottom rows, $xa = 1$ which can't happen and have $a = 0$. Therefore, you can't write this matrix in the form LU. It has no LU factorization. This is what we mean above by saying the method lacks generality.

8.2 Finding An LU Factorization By Inspection

Which matrices have an LU factorization? It turns out it is those whose row reduced echelon form can be achieved without switching rows and which only involve row operations of type 3 in which row j is replaced with a multiple of row i added to row j for $i < j$.

Example 8.2.1 *Find an LU factorization of* $A = \begin{pmatrix} 1 & 2 & 0 & 2 \\ 1 & 3 & 2 & 1 \\ 2 & 3 & 4 & 0 \end{pmatrix}$.

One way to find the LU factorization is to simply look for it directly. You need

$$\begin{pmatrix} 1 & 2 & 0 & 2 \\ 1 & 3 & 2 & 1 \\ 2 & 3 & 4 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ x & 1 & 0 \\ y & z & 1 \end{pmatrix} \begin{pmatrix} a & d & h & j \\ 0 & b & e & i \\ 0 & 0 & c & f \end{pmatrix}.$$

Then multiplying these you get

$$\begin{pmatrix} a & d & h & j \\ xa & xd+b & xh+e & xj+i \\ ya & yd+zb & yh+ze+c & yj+iz+f \end{pmatrix}$$

and so you can now tell what the various quantities equal. From the first column, you need $a = 1, x = 1, y = 2$. Now go to the second column. You need $d = 2, xd + b = 3$ so $b = 1, yd + zb = 3$ so $z = -1$. From the third column, $h = 0, e = 2, c = 6$. Now from the fourth column, $j = 2, i = -1, f = -5$. Therefore, an LU factorization is

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 & 2 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 6 & -5 \end{pmatrix}.$$

You can check whether you got it right by simply multiplying these two.

8.3 Using Multipliers To Find An LU Factorization

There is also a convenient procedure for finding an LU factorization. It turns out that it is only necessary to keep track of the **multipliers** which are used to row reduce to upper triangular form. This procedure is described in the following examples.

Example 8.3.1 *Find an LU factorization for* $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & -4 \\ 1 & 5 & 2 \end{pmatrix}$

Write the matrix next to the identity matrix as shown.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & -4 \\ 1 & 5 & 2 \end{pmatrix}.$$

The process involves doing row operations to the matrix on the right while simultaneously updating successive columns of the matrix on the left. First take -2 times the first row and add to the second in the matrix on the right.

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -10 \\ 1 & 5 & 2 \end{pmatrix}$$

Note the way we updated the matrix on the left. We put a 2 in the second entry of the first column because we used -2 times the first row added to the second row. Now replace the

third row in the matrix on the right by -1 times the first row added to the third. Thus the next step is

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -10 \\ 0 & 3 & -1 \end{pmatrix}$$

Finally, we will add the second row to the bottom row and make the following changes

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -10 \\ 0 & 0 & -11 \end{pmatrix}.$$

At this point, we stop because the matrix on the right is upper triangular. An LU factorization is the above.

The justification for this gimmick will be given later.

Example 8.3.2 *Find an LU factorization for* $A = \begin{pmatrix} 1 & 2 & 1 & 2 & 1 \\ 2 & 0 & 2 & 1 & 1 \\ 2 & 3 & 1 & 3 & 2 \\ 1 & 0 & 1 & 1 & 2 \end{pmatrix}$.

We will use the same procedure as above. However, this time we will do everything for one column at a time. First multiply the first row by (-1) and then add to the last row. Next take (-2) times the first and add to the second and then (-2) times the first and add to the third.

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 & 2 & 1 \\ 0 & -4 & 0 & -3 & -1 \\ 0 & -1 & -1 & -1 & 0 \\ 0 & -2 & 0 & -1 & 1 \end{pmatrix}.$$

This finishes the first column of L and the first column of U. Now take $-(1/4)$ times the second row in the matrix on the right and add to the third followed by $-(1/2)$ times the second added to the last.

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 2 & 1/4 & 1 & 0 \\ 1 & 1/2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 1 & 2 & 1 \\ 0 & -4 & 0 & -3 & -1 \\ 0 & 0 & -1 & -1/4 & 1/4 \\ 0 & 0 & 0 & 1/2 & 3/2 \end{pmatrix}$$

This finishes the second column of L as well as the second column of U. Since the matrix on the right is upper triangular, stop. The LU factorization has now been obtained. This technique is called Dolittle's method.

This process is entirely typical of the general case. The matrix U is just the first upper triangular matrix you come to in your quest for the row reduced echelon form using only the row operation which involves replacing a row by itself added to a multiple of another row. The matrix, L is what you get by updating the identity matrix as illustrated above.

You should note that for a square matrix, the number of row operations necessary to reduce to LU form is about half the number needed to place the matrix in row reduced echelon form. This is why an LU factorization is of interest in solving systems of equations.

8.4 Solving Systems Using The LU Factorization

The reason people care about the LU factorization is it allows the quick solution of systems of equations. Here is an example.

Example 8.4.1 *Suppose you want to find the solutions to* $\begin{pmatrix} 1 & 2 & 3 & 2 \\ 4 & 3 & 1 & 1 \\ 1 & 2 & 3 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

Of course one way is to write the augmented matrix and grind away. However, this involves more row operations than the computation of the LU factorization and it turns out that the LU factorization can give the solution quickly. Here is how. The following is an LU factorization for the matrix.

$$\begin{pmatrix} 1 & 2 & 3 & 2 \\ 4 & 3 & 1 & 1 \\ 1 & 2 & 3 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{pmatrix}.$$

Let $U\mathbf{x} = \mathbf{y}$ and consider $L\mathbf{y} = \mathbf{b}$ where in this case, $\mathbf{b} = (1, 2, 3)^T$. Thus

$$\begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

which yields very quickly that $\mathbf{y} = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}$. Now you can find $\mathbf{x}$ by solving $U\mathbf{x} = \mathbf{y}$. Thus in this case,

$$\begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}$$

which yields

$$\mathbf{x} = \begin{pmatrix} -\frac{3}{5} + \frac{7}{5}t \\ \frac{9}{5} - \frac{11}{5}t \\ t \\ -1 \end{pmatrix}, t \in \mathbb{R}.$$

8.5 Justification For The Multiplier Method

Why does the multiplier method work for finding the LU factorization? Suppose A is a matrix which has the property that the row reduced echelon form for A may be achieved using only the row operations which involve replacing a row with itself added to a multiple of another row. It is not ever necessary to switch rows. Thus every row which is replaced using this row operation in obtaining the echelon form may be modified by using a row which is above it. Furthermore, in the multiplier method for finding the LU factorization, we zero out the elements below the pivot element in first column and then the next and so on when scanning from the left. In terms of elementary matrices, this means the row operations used to reduce A to upper triangular form correspond to multiplication on the left by lower triangular matrices having all ones down the main diagonal.and the sequence of elementary matrices which row reduces A has the property that in scanning the list of

elementary matrices from the right to the left, this list consists of several matrices which involve only changes from the identity in the first column, then several which involve only changes from the identity in the second column and so forth. More precisely, $E_p \cdots E_1 A = U$ where U is upper triangular, each E_i is a lower triangular elementary matrix having all ones down the main diagonal, for some r_i, each of $E_{r_1} \cdots E_1$ differs from the identity only in the first column, each of $E_{r_2} \cdots E_{r_1+1}$ differs from the identity only in the second column and so forth. Therefore, $A = \overbrace{E_1^{-1} \cdots E_{p-1}^{-1} E_p^{-1}}^{\text{Will be } L} U$. You multiply the inverses in the reverse order. Now each of the E_i^{-1} is also lower triangular with 1 down the main diagonal. Therefore their product has this property. Recall also that if E_i equals the identity matrix except for having an a in the j^{th} column somewhere below the main diagonal, E_i^{-1} is obtained by replacing the a in E_i with $-a$ thus explaining why we replace with -1 times the multiplier in computing L. In the case where A is a $3 \times m$ matrix, $E_1^{-1} \cdots E_{p-1}^{-1} E_p^{-1}$ is of the form

$$\begin{pmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ b & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & c & 1 \end{pmatrix}.$$

Note that scanning from left to right, the first two in the product involve changes in the identity only in the first column while in the third matrix, the change is only in the second. If we had zeroed out the elements of the first column in a different order, we would have obtained.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ b & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & c & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & c & 1 \end{pmatrix}$$

However, it is important to be working from the left to the right, one column at a time.

A similar observation holds in any dimension. Multiplying the elementary matrices which involve a change only in the j^{th} column you obtain A equal to an upper triangular, $n \times m$ matrix, U multiplied on its left by a sequence of lower triangular matrices which is of the following form, the a_{ij} being negatives of multipliers used in row reducing A to an upper triangular matrix.

$$\begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ a_{11} & 1 & & & & \vdots \\ \vdots & 0 & \ddots & & & \vdots \\ \vdots & \vdots & 0 & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & 0 \\ a_{1,n-1} & 0 & 0 & \cdots & \cdots & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & & & & \vdots \\ \vdots & a_{21} & \ddots & & & \vdots \\ \vdots & \vdots & 0 & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & 0 \\ 0 & a_{2,n-2} & 0 & \cdots & \cdots & 1 \end{pmatrix} \cdots$$

$$\cdots \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & & & & \vdots \\ \vdots & 0 & \ddots & & & \vdots \\ \vdots & \vdots & 0 & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & 1 & 0 \\ 0 & 0 & 0 & \cdots & a_{n,n-1} & 1 \end{pmatrix}$$

From the way we multiply matrices, this product equals

$$\begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ a_{11} & 1 & & & & \vdots \\ a_{12} & a_{21} & \ddots & & & \vdots \\ \vdots & a_{22} & a_{31} & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & 1 & 0 \\ a_{1,n-1} & a_{2,n-2} & a_{3,n-3} & \cdots & a_{n,n-1} & 1 \end{pmatrix}$$

Notice how the end result of the matrix multiplication made no change in the a_{ij}. It just filled in the empty spaces with the a_{ij} which occurred in one of the matrices in the product. This is why, in computing L, it is sufficient to begin with the left column and work column by column toward the right, replacing entries with the negative of the multiplier used in the row operation which produces a zero in that entry.

8.6 The PLU Factorization

As indicated above, some matrices don't have an LU factorization. Here is an example.

$$M = \begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & 2 & 3 & 0 \\ 4 & 3 & 1 & 1 \end{pmatrix} \tag{8.1}$$

In this case, there is another factorization which is useful called a PLU factorization. Here P is a permutation matrix.

Example 8.6.1 *Find a PLU factorization for the above matrix in 8.1.*

Proceed as before trying to find the row echelon form of the matrix. First add -1 times the first row to the second row and then add -4 times the first to the third. This yields

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 4 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & 0 & 0 & -2 \\ 0 & -5 & -11 & -7 \end{pmatrix}$$

There is no way to do only row operations involving replacing a row with itself added to a multiple of another row to the matrix on the right in such a way as to obtain an upper triangular matrix. Therefore, consider the original matrix with the bottom two rows switched.

$$\begin{aligned} M' &= \begin{pmatrix} 1 & 2 & 3 & 2 \\ 4 & 3 & 1 & 1 \\ 1 & 2 & 3 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & 2 & 3 & 0 \\ 4 & 3 & 1 & 1 \end{pmatrix} \\ &= PM \end{aligned}$$

Now try again with this matrix. First take -1 times the first row and add to the bottom row and then take -4 times the first row and add to the second row. This yields

$$\begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{pmatrix}$$

The matrix on the right is upper triangular and so the LU factorization of the matrix, M' has been obtained above.

Thus $M' = PM = LU$ where L and U are given above. Notice that $P^2 = I$ and therefore, $M = P^2M = PLU$ and so

$$\begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & 2 & 3 & 0 \\ 4 & 3 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{pmatrix}$$

This process can always be followed and so there always exists a PLU factorization of a given matrix even though there isn't always an LU factorization.

Example 8.6.2 *Use the PLU factorization of* $M \equiv \begin{pmatrix} 1 & 2 & 3 & 2 \\ 1 & 2 & 3 & 0 \\ 4 & 3 & 1 & 1 \end{pmatrix}$ *to solve the system* $M\mathbf{x} = \mathbf{b}$ *where* $\mathbf{b} = (1, 2, 3)^T$.

Let $U\mathbf{x} = \mathbf{y}$ and consider $PL\mathbf{y} = \mathbf{b}$. In other words, solve,

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Multiplying both sides by P gives

$$\begin{pmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}$$

and so

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}.$$

Now $U\mathbf{x} = \mathbf{y}$ and so it only remains to solve

$$\begin{pmatrix} 1 & 2 & 3 & 2 \\ 0 & -5 & -11 & -7 \\ 0 & 0 & 0 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

which yields

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} \frac{1}{5} + \frac{7}{5}t \\ \frac{9}{10} - \frac{11}{5}t \\ t \\ -\frac{1}{2} \end{pmatrix} : t \in \mathbb{R}.$$

8.7 The QR Factorization

As pointed out above, the LU factorization is not a mathematically respectable thing because it does not always exist. There is another factorization which does always exist. Much more can be said about it than I will say here. I will only deal with real matrices and so the dot product will be the usual real dot product.

Definition 8.7.1 *An $n \times n$ real matrix Q is called an orthogonal matrix if*

$$QQ^T = Q^T Q = I.$$

Thus an orthogonal matrix is one whose inverse is equal to its transpose.

First note that if a matrix is orthogonal this says

$$\sum_j Q_{ij}^T Q_{jk} = \sum_j Q_{ji} Q_{jk} = \delta_{ik}$$

Thus

$$|Q\mathbf{x}|^2 = \sum_i \left(\sum_j Q_{ij} x_j \right)^2 = \sum_i \sum_r \sum_s Q_{is} x_s Q_{ir} x_r$$

$$= \sum_i \sum_r \sum_s Q_{is} Q_{ir} x_s x_r = \sum_r \sum_s \sum_i Q_{is} Q_{ir} x_s x_r$$

$$= \sum_r \sum_s \delta_{sr} x_s x_r = \sum_r x_r^2 = |\mathbf{x}|^2$$

This shows that orthogonal transformations preserve distances. You can show that if you have a matrix which does preserve distances, then it must be orthogonal also.

Example 8.7.2 *One of the most important examples of an orthogonal matrix is the so called Householder matrix. You have $\mathbf{v}$ a unit vector and you form the matrix,*

$$I - 2\mathbf{v}\mathbf{v}^T$$

This is an orthogonal matrix which is also symmetric. To see this, you use the rules of matrix operations.

$$\begin{aligned} \left(I - 2\mathbf{v}\mathbf{v}^T\right)^T &= I^T - \left(2\mathbf{v}\mathbf{v}^T\right)^T \\ &= I - 2\mathbf{v}\mathbf{v}^T \end{aligned}$$

so it is symmetric. Now to show it is orthogonal,

$$\begin{aligned} \left(I - 2\mathbf{v}\mathbf{v}^T\right)\left(I - 2\mathbf{v}\mathbf{v}^T\right) &= I - 2\mathbf{v}\mathbf{v}^T - 2\mathbf{v}\mathbf{v}^T + 4\mathbf{v}\mathbf{v}^T\mathbf{v}\mathbf{v}^T \\ &= I - 4\mathbf{v}\mathbf{v}^T + 4\mathbf{v}\mathbf{v}^T = I \end{aligned}$$

because $\mathbf{v}^T\mathbf{v} = \mathbf{v} \cdot \mathbf{v} = |\mathbf{v}|^2 = 1$. Therefore, this is an example of an orthogonal matrix.

Consider the following problem.

Problem 8.7.3 *Given two vectors $\mathbf{x}, \mathbf{y}$ such that $|\mathbf{x}| = |\mathbf{y}| \neq 0$ but $\mathbf{x} \neq \mathbf{y}$ and you want an orthogonal matrix, Q such that $Q\mathbf{x} = \mathbf{y}$ and $Q\mathbf{y} = \mathbf{x}$. The thing which works is the Householder matrix*

$$Q \equiv I - 2\frac{\mathbf{x} - \mathbf{y}}{|\mathbf{x} - \mathbf{y}|^2} (\mathbf{x} - \mathbf{y})^T$$

Here is why this works.

$$\begin{aligned} Q(\mathbf{x} - \mathbf{y}) &= (\mathbf{x} - \mathbf{y}) - 2\frac{\mathbf{x} - \mathbf{y}}{|\mathbf{x} - \mathbf{y}|^2} (\mathbf{x} - \mathbf{y})^T (\mathbf{x} - \mathbf{y}) \\ &= (\mathbf{x} - \mathbf{y}) - 2\frac{\mathbf{x} - \mathbf{y}}{|\mathbf{x} - \mathbf{y}|^2} |\mathbf{x} - \mathbf{y}|^2 = \mathbf{y} - \mathbf{x} \end{aligned}$$

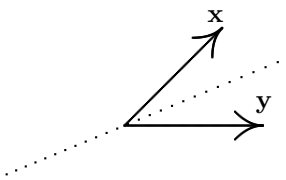
$$\begin{aligned} Q(\mathbf{x}+\mathbf{y}) &= (\mathbf{x}+\mathbf{y}) - 2\frac{\mathbf{x}-\mathbf{y}}{|\mathbf{x}-\mathbf{y}|^2}(\mathbf{x}-\mathbf{y})^T(\mathbf{x}+\mathbf{y}) \\ &= (\mathbf{x}+\mathbf{y}) - 2\frac{\mathbf{x}-\mathbf{y}}{|\mathbf{x}-\mathbf{y}|^2}((\mathbf{x}-\mathbf{y})\cdot(\mathbf{x}+\mathbf{y})) \\ &= (\mathbf{x}+\mathbf{y}) - 2\frac{\mathbf{x}-\mathbf{y}}{|\mathbf{x}-\mathbf{y}|^2}\left(|\mathbf{x}|^2-|\mathbf{y}|^2\right) = \mathbf{x}+\mathbf{y} \end{aligned}$$

Hence

$$\begin{aligned} Q\mathbf{x}+Q\mathbf{y} &= \mathbf{x}+\mathbf{y} \\ Q\mathbf{x}-Q\mathbf{y} &= \mathbf{y}-\mathbf{x} \end{aligned}$$

Adding these equations, $2Q\mathbf{x}=2\mathbf{y}$ *and subtracting them yields* $2Q\mathbf{y}=2\mathbf{x}$.

A picture of the geometric significance follows.



The orthogonal matrix Q reflects across the dotted line taking $\mathbf{x}$ to $\mathbf{y}$ and $\mathbf{y}$ to $\mathbf{x}$.

Definition 8.7.4 *Let A be an $m\times n$ matrix. Then a QR factorization of A consists of two matrices, Q orthogonal and R upper triangular or in other words equal to zero below the main diagonal such that $A=QR$.*

With the solution to this simple problem, here is how to obtain a QR factorization for any matrix A. Let

$$A=(\mathbf{a}_1,\mathbf{a}_2,\cdots,\mathbf{a}_n)$$

where the $\mathbf{a}_i$ are the columns. If $\mathbf{a}_1=\mathbf{0}$, let $Q_1=I$. If $\mathbf{a}_1\neq\mathbf{0}$, let

$$\mathbf{b}\equiv\begin{pmatrix}|\mathbf{a}_1|\\0\\\vdots\\0\end{pmatrix}$$

and form the Householder matrix,

$$Q_1\equiv I-2\frac{(\mathbf{a}_1-\mathbf{b})}{|\mathbf{a}_1-\mathbf{b}|^2}(\mathbf{a}_1-\mathbf{b})^T$$

As in the above problem $Q_1\mathbf{a}_1=\mathbf{b}$ and so

$$Q_1A=\begin{pmatrix}|\mathbf{a}_1| & *\\ \mathbf{0} & A_2\end{pmatrix}$$

where A_2 is a $m-1\times n-1$ matrix. Now find in the same way as was just done a $n-1\times n-1$ matrix $\widehat{Q}_2$ such that

$$\widehat{Q}_2A_2=\begin{pmatrix}* & *\\ \mathbf{0} & A_3\end{pmatrix}$$

Let

$$Q_2 \equiv \begin{pmatrix} 1 & 0 \\ \mathbf{0} & \widehat{Q}_2 \end{pmatrix}.$$

Then

$$Q_2 Q_1 A = \begin{pmatrix} 1 & 0 \\ \mathbf{0} & \widehat{Q}_2 \end{pmatrix} \begin{pmatrix} |\mathbf{a}_1| & * \\ \mathbf{0} & A_2 \end{pmatrix}$$

$$= \begin{pmatrix} |\mathbf{a}_1| & * & * \\ \vdots & * & * \\ 0 & \mathbf{0} & A_3 \end{pmatrix}$$

Continuing this way untill the result is upper triangular, you get a sequence of orthogonal matrices $Q_p Q_{p-1} \cdots Q_1$ such that

$$Q_p Q_{p-1} \cdots Q_1 A = R \tag{8.2}$$

where R is upper triangular.

Now if Q_1 and Q_2 are orthogonal, then from properties of matrix multiplication,

$$Q_1 Q_2 \left(Q_1 Q_2\right)^T = Q_1 Q_2 Q_2^T Q_1^T = Q_1 I Q_1^T = I$$

and similarly

$$\left(Q_1 Q_2\right)^T Q_1 Q_2 = I.$$

Thus the product of orthogonal matrices is orthogonal. Also the transpose of an orthogonal matrix is orthogonal directly from the definition. Therefore, from 8.2

$$A = \left(Q_p Q_{p-1} \cdots Q_1\right)^T R \equiv QR.$$

This proves the following theorem.

Theorem 8.7.5 *Let A be any real $m \times n$ matrix. Then there exists an orthogonal matrix, Q and an upper triangular matrix R such that*

$$A = QR$$

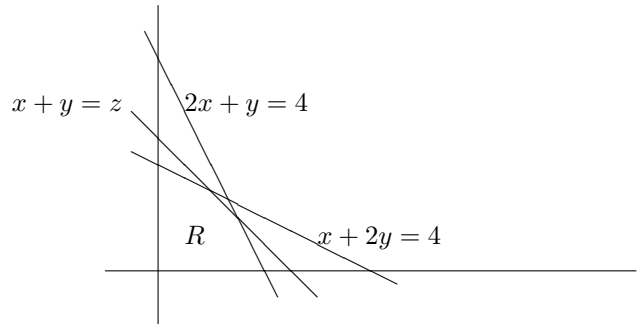
and this factorization can be accomplished in a systematic manner.

Linear Programming

9.1 Simple Geometric Considerations

One of the most important uses of row operations is in solving linear program problems which involve maximizing a linear function subject to inequality constraints determined from linear equations. Here is an example. A certain hamburger store has 9000 hamburger patties to use in one week and a limitless supply of special sauce, lettuce, tomatoes, onions, and buns. They sell two types of hamburgers, the big stack and the basic burger. It has also been determined that the employees cannot prepare more than 9000 of either type in one week. The big stack, popular with the teen agers from the local high school, involves two patties, lots of delicious sauce, condiments galore, and a divider between the two patties. The basic burger, very popular with children, involves only one patty and some pickles and ketchup. Demand for the basic burger is twice what it is for the big stack. What is the maximum number of hamburgers which could be sold in one week given the above limitations?

Let x be the number of basic burgers and y the number of big stacks which could be sold in a week. Thus it is desired to maximize $z = x + y$ subject to the above constraints. The total number of patties is 9000 and so the number of patty used is $x+2y$. This number must satisfy $x + 2y \leq 9000$ because there are only 9000 patty available. Because of the limitation on the number the employees can prepare and the demand, it follows $2x + y \leq 9000$. You never sell a negative number of hamburgers and so $x, y \geq 0$. In simpler terms the problem reduces to maximizing $z = x + y$ subject to the two constraints, $x + 2y \leq 9000$ and $2x + y \leq 9000$. This problem is pretty easy to solve geometrically. Consider the following picture in which R labels the region described by the above inequalities and the line $z = x+y$ is shown for a particular value of z.



As you make z larger this line moves away from the origin, always having the same slope

and the desired solution would consist of a point in the region, R which makes z as large as possible or equivalently one for which the line is as far as possible from the origin. Clearly this point is the point of intersection of the two lines, $(3000, 3000)$ and so the maximum value of the given function is 6000. Of course this type of procedure is fine for a situation in which there are only two variables but what about a similar problem in which there are very many variables. In reality, this hamburger store makes many more types of burgers than those two and there are many considerations other than demand and available patty. Each will likely give you a constraint which must be considered in order to solve a more realistic problem and the end result will likely be a problem in many dimensions, probably many more than three so your ability to draw a picture will get you nowhere for such a problem. Another method is needed. This method is the topic of this section. I will illustrate with this particular problem. Let $x_1 = x$ and $y = x_2$. Also let x_3 and x_4 be nonnegative variables such that

$$x_1 + 2x_2 + x_3 = 9000,\ 2x_1 + x_2 + x_4 = 9000.$$

To say that x_3 and x_4 are nonnegative is the same as saying $x_1 + 2x_2 \leq 9000$ and $2x_1 + x_2 \leq 9000$ and these variables are called slack variables at this point. They are called this because they "take up the slack". I will discuss these more later. First a general situation is considered.

9.2 The Simplex Tableau

Here is some notation.

Definition 9.2.1 *Let* $\mathbf{x}, \mathbf{y}$ *be vectors in* $\mathbb{R}^q$. *Then* $\mathbf{x} \leq \mathbf{y}$ *means for each* $i, x_i \leq y_i$.

The problem is as follows:

Let A be an $m \times (m+n)$ real matrix of rank m. It is desired to find $\mathbf{x} \in \mathbb{R}^{n+m}$ such that $\mathbf{x}$ satisfies the constraints,

$$\mathbf{x} \geq \mathbf{0},\ A\mathbf{x} = \mathbf{b} \tag{9.1}$$

and out of all such $\mathbf{x}$,

$$z \equiv \sum_{i=1}^{m+n} c_i x_i$$

is as large (or small) as possible. This is usually referred to as maximizing or minimizing z subject to the above constraints. First I will consider the constraints.

Let $A = \begin{pmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_{n+m} \end{pmatrix}$. First you find a vector, $\mathbf{x}^0 \geq \mathbf{0}$, $A\mathbf{x}^0 = \mathbf{b}$ such that n of the components of this vector equal 0. Letting $i_1, \cdots, i_n$ be the positions of $\mathbf{x}^0$ for which $x_i^0 = 0$, suppose also that $\{\mathbf{a}_{j_1}, \cdots, \mathbf{a}_{j_m}\}$ is linearly independent for j_i the other positions of $\mathbf{x}^0$. Geometrically, this means that $\mathbf{x}^0$ is a corner of the feasible region, those $\mathbf{x}$ which satisfy the constraints. This is called a basic feasible solution. Also define

$$\begin{aligned} \mathbf{c}_B &\equiv (c_{j_1}, \cdots, c_{j_m}),\ \mathbf{c}_F \equiv (c_{i_1}, \cdots, c_{i_n}) \\ \mathbf{x}_B &\equiv (x_{j_1}, \cdots, x_{j_m}),\ \mathbf{x}_F \equiv (x_{i_1}, \cdots, x_{i_n}). \end{aligned}$$

and

$$z^0 \equiv z\left(\mathbf{x}^0\right) = \begin{pmatrix} \mathbf{c}_B & \mathbf{c}_F \end{pmatrix} \begin{pmatrix} \mathbf{x}_B^0 \\ \mathbf{x}_F^0 \end{pmatrix} = \mathbf{c}_B \mathbf{x}_B^0$$

since $\mathbf{x}_F^0 = \mathbf{0}$. The variables which are the components of the vector $\mathbf{x}_B$ are called the **basic variables** and the variables which are the entries of $\mathbf{x}_F$ are called the **free variables.** You

set $\mathbf{x}_F = \mathbf{0}$. Now $\left(\mathbf{x}^0, z^0\right)^T$ is a solution to

$$\begin{pmatrix} A & \mathbf{0} \\ -\mathbf{c} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ z \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ 0 \end{pmatrix}$$

along with the constraints $\mathbf{x} \geq \mathbf{0}$. Writing the above in augmented matrix form yields

$$\begin{pmatrix} A & \mathbf{0} & \mathbf{b} \\ -\mathbf{c} & 1 & 0 \end{pmatrix} \tag{9.2}$$

Permute the columns and variables on the left if necessary to write the above in the form

$$\begin{pmatrix} B & F & \mathbf{0} \\ -\mathbf{c}_B & -\mathbf{c}_F & 1 \end{pmatrix} \begin{pmatrix} \mathbf{x}_B \\ \mathbf{x}_F \\ z \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ 0 \end{pmatrix} \tag{9.3}$$

or equivalently in the augmented matrix form keeping track of the variables on the bottom as

$$\begin{pmatrix} B & F & \mathbf{0} & \mathbf{b} \\ -\mathbf{c}_B & -\mathbf{c}_F & 1 & 0 \\ \mathbf{x}_B & \mathbf{x}_F & 0 & 0 \end{pmatrix}. \tag{9.4}$$

Here B pertains to the variables $x_{i_1}, \cdots, x_{j_m}$ and is an $m \times m$ matrix with linearly independent columns, $\{\mathbf{a}_{j_1}, \cdots, \mathbf{a}_{j_m}\}$, and F is an $m \times n$ matrix. Now it is assumed that

$$\begin{pmatrix} B & F \end{pmatrix} \begin{pmatrix} \mathbf{x}_B^0 \\ \mathbf{x}_F^0 \end{pmatrix} = \begin{pmatrix} B & F \end{pmatrix} \begin{pmatrix} \mathbf{x}_B^0 \\ \mathbf{0} \end{pmatrix} = B\mathbf{x}_B^0 = \mathbf{b}$$

and since B is assumed to have rank m, it follows

$$\mathbf{x}_B^0 = B^{-1}\mathbf{b} \geq \mathbf{0}. \tag{9.5}$$

This is very important to observe. $B^{-1}\mathbf{b} \geq \mathbf{0}$!

Do row operations on the top part of the matrix,

$$\begin{pmatrix} B & F & \mathbf{0} & \mathbf{b} \\ -\mathbf{c}_B & -\mathbf{c}_F & 1 & 0 \end{pmatrix} \tag{9.6}$$

and obtain its row reduced echelon form. Then after these row operations the above becomes

$$\begin{pmatrix} I & B^{-1}F & \mathbf{0} & B^{-1}\mathbf{b} \\ -\mathbf{c}_B & -\mathbf{c}_F & 1 & 0 \end{pmatrix}. \tag{9.7}$$

where $B^{-1}\mathbf{b} \geq 0$. Next do another row operation in order to get a $\mathbf{0}$ where you see a $-\mathbf{c}_B$. Thus

$$\begin{pmatrix} I & B^{-1}F & \mathbf{0} & B^{-1}\mathbf{b} \\ \mathbf{0} & \mathbf{c}_B B^{-1}F' - \mathbf{c}_F & 1 & \mathbf{c}_B B^{-1}\mathbf{b} \end{pmatrix} \tag{9.8}$$

$$= \begin{pmatrix} I & B^{-1}F & \mathbf{0} & B^{-1}\mathbf{b} \\ \mathbf{0} & \mathbf{c}_B B^{-1}F' - \mathbf{c}_F & 1 & \mathbf{c}_B \mathbf{x}_B^0 \end{pmatrix}$$

$$= \begin{pmatrix} I & B^{-1}F & \mathbf{0} & B^{-1}\mathbf{b} \\ \mathbf{0} & \mathbf{c}_B B^{-1}F - \mathbf{c}_F & 1 & z^0 \end{pmatrix} \tag{9.9}$$

The reason there is a z^0 on the bottom right corner is that $\mathbf{x}_F = 0$ and $\left(\mathbf{x}_B^0, \mathbf{x}_F^0, z^0\right)^T$ is a solution of the system of equations represented by the above augmented matrix because it is

a solution to the system of equations corresponding to the system of equations represented by 9.6 and row operations leave solution sets unchanged. Note how attractive this is. The z_0 is the value of z at the point $\mathbf{x}^0$. The augmented matrix of 9.9 is called the simplex tableau and it is the beginning point for the simplex algorithm to be described a little later. It is very convenient to express the simplex tableau in the above form in which the variables are possibly permuted in order to have $\begin{pmatrix} I \\ \mathbf{0} \end{pmatrix}$ on the left side. However, as far as the simplex algorithm is concerned it is not necessary to be permuting the variables in this manner. Starting with 9.9 you could permute the variables and columns to obtain an augmented matrix in which the variables are in their original order. What is really required for the simplex tableau?

It is an augmented $m+1 \times m+n+2$ matrix which represents a system of equations which has the same set of solutions, $(\mathbf{x},z)^T$ as the system whose augmented matrix is

$$\begin{pmatrix} A & \mathbf{0} & \mathbf{b} \\ -\mathbf{c} & 1 & 0 \end{pmatrix}$$

(Possibly the variables for $\mathbf{x}$ are taken in another order.) There are m linearly independent columns in the first $m+n$ columns for which there is only one nonzero entry, a 1 in one of the first m rows, the "simple columns", the other first $m+n$ columns being the "nonsimple columns". As in the above, the variables corresponding to the simple columns are $\mathbf{x}_B$, the basic variables and those corresponding to the nonsimple columns are $\mathbf{x}_F$, the free variables. Also, the top m entries of the last column on the right are nonnegative. This is the description of a simplex tableau.

In a simplex tableau it is easy to spot a basic feasible solution. You can see one quickly by setting the variables, $\mathbf{x}_F$ corresponding to the nonsimple columns equal to zero. Then the other variables, corresponding to the simple columns are each equal to a nonnegative entry in the far right column. Lets call this an "obvious basic feasible solution". If a solution is obtained by setting the variables corresponding to the nonsimple columns equal to zero and the variables corresponding to the simple columns equal to zero this will be referred to as an "obvious" solution. Lets also call the first $m+n$ entries in the bottom row the "bottom left row". In a simplex tableau, the entry in the bottom right corner gives the value of the variable being maximized or minimized when the obvious basic feasible solution is chosen.

The following is a special case of the general theory presented above and shows how such a special case can be fit into the above framework. The following example is rather typical of the sorts of problems considered. It involves inequality constraints instead of $A\mathbf{x} = \mathbf{b}$. This is handled by adding in "slack variables" as explained below.

Example 9.2.2 *Consider $z = x_1 - x_2$ subject to the constraints, $x_1+2x_2 \leq 10, x_1+2x_2 \geq 2$, and $2x_1 + x_2 \leq 6, x_i \geq 0$. Find a simplex tableau for a problem of the form $\mathbf{x} \geq \mathbf{0}, A\mathbf{x} = \mathbf{b}$ which is equivalent to the above problem.*

You add in slack variables. These are positive variables, one for each of the first three constraints, which change the first three inequalities into equations. Thus the first three inequalities become $x_1+2x_2+x_3 = 10, x_1+2x_2-x_4 = 2$, and $2x_1+x_2+x_5 = 6, x_1, x_2, x_3, x_4, x_5 \geq 0$. Now it is necessary to find a basic feasible solution. You mainly need to find a positive solution to the equations,

$$\begin{array}{c} x_1 + 2x_2 + x_3 = 10 \\ x_1 + 2x_2 - x_4 = 2 \\ 2x_1 + x_2 + x_5 = 6 \end{array}.$$

the solution set for the above system is given by

$$x_2 = \frac{2}{3}x_4 - \frac{2}{3} + \frac{1}{3}x_5, x_1 = -\frac{1}{3}x_4 + \frac{10}{3} - \frac{2}{3}x_5, x_3 = -x_4 + 8.$$

An easy way to get a basic feasible solution is to let $x_4 = 8$ and $x_5 = 1$. Then a feasible solution is

$$(x_1, x_2, x_3, x_4, x_5) = (0, 5, 0, 8, 1)\,.$$

It follows $z^0 = -5$ and the matrix 9.2, $\begin{pmatrix} A & \mathbf{0} & \mathbf{b} \\ -\mathbf{c} & 1 & 0 \end{pmatrix}$ with the variables kept track of on the bottom is

$$\begin{pmatrix} 1 & 2 & 1 & 0 & 0 & 0 & 10 \\ 1 & 2 & 0 & -1 & 0 & 0 & 2 \\ 2 & 1 & 0 & 0 & 1 & 0 & 6 \\ -1 & 1 & 0 & 0 & 0 & 1 & 0 \\ x_1 & x_2 & x_3 & x_4 & x_5 & 0 & 0 \end{pmatrix}$$

and the first thing to do is to permute the columns so that the list of variables on the bottom will have x_1 and x_3 at the end.

$$\begin{pmatrix} 2 & 0 & 0 & 1 & 1 & 0 & 10 \\ 2 & -1 & 0 & 1 & 0 & 0 & 2 \\ 1 & 0 & 1 & 2 & 0 & 0 & 6 \\ 1 & 0 & 0 & -1 & 0 & 1 & 0 \\ x_2 & x_4 & x_5 & x_1 & x_3 & 0 & 0 \end{pmatrix}$$

Next, as described above, take the row reduced echelon form of the top three lines of the above matrix. This yields

$$\begin{pmatrix} 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 5 \\ 0 & 1 & 0 & 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & \frac{3}{2} & -\frac{1}{2} & 0 & 1 \end{pmatrix}.$$

Now do row operations to

$$\begin{pmatrix} 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 5 \\ 0 & 1 & 0 & 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & \frac{3}{2} & -\frac{1}{2} & 0 & 1 \\ 1 & 0 & 0 & -1 & 0 & 1 & 0 \end{pmatrix}$$

to finally obtain

$$\begin{pmatrix} 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 5 \\ 0 & 1 & 0 & 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & \frac{3}{2} & -\frac{1}{2} & 0 & 1 \\ 0 & 0 & 0 & -\frac{3}{2} & -\frac{1}{2} & 1 & -5 \end{pmatrix}$$

and this is a simplex tableau. The variables are $x_2, x_4, x_5, x_1, x_3, z$.

It isn't as hard as it may appear from the above. Lets not permute the variables and simply find an acceptable simplex tableau as described above.

Example 9.2.3 *Consider $z = x_1 - x_2$ subject to the constraints, $x_1 + 2x_2 \leq 10, x_1 + 2x_2 \geq 2$, and $2x_1 + x_2 \leq 6, x_i \geq 0$. Find a simplex tableau.*

Adding in slack variables, an augmented matrix which is descriptive of the constraints is

$$\begin{pmatrix} 1 & 2 & 1 & 0 & 0 & 10 \\ 1 & 2 & 0 & -1 & 0 & 6 \\ 2 & 1 & 0 & 0 & 1 & 6 \end{pmatrix}$$

The obvious solution is not feasible because of that -1 in the fourth column. Consider the second column and select the 2 as a pivot to zero out that which is above and below the 2. This is because that 2 satisfies the criterion for being chosen as a pivot.

$$\begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 4 \\ \frac{1}{2} & 1 & 0 & -\frac{1}{2} & 0 & 3 \\ \frac{3}{2} & 0 & 0 & \frac{1}{2} & 1 & 3 \end{pmatrix}$$

This one is good. The obvious solution is now feasible. You can now assemble the simplex tableau. The first step is to include a column and row for z. This yields

$$\begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 4 \\ \frac{1}{2} & 1 & 0 & -\frac{1}{2} & 0 & 0 & 3 \\ \frac{3}{2} & 0 & 0 & \frac{1}{2} & 1 & 0 & 3 \\ -1 & 0 & 1 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Now you need to get zeros in the right places so the simple columns will be preserved as simple columns. This means you need to zero out the 1 in the third column on the bottom. A simplex tableau is now

$$\begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 4 \\ \frac{1}{2} & 1 & 0 & -\frac{1}{2} & 0 & 0 & 3 \\ \frac{3}{2} & 0 & 0 & \frac{1}{2} & 1 & 0 & 3 \\ -1 & 0 & 0 & -1 & 0 & 1 & -4 \end{pmatrix}.$$

Note it is not the same one obtained earlier. There is no reason a simplex tableau should be unique. In fact, it follows from the above general description that you have one for each basic feasible point of the region determined by the constraints.

9.3 The Simplex Algorithm

9.3.1 Maximums

The simplex algorithm takes you from one basic feasible solution to another while maximizing or minimizing the function you are trying to maximize or minimize. Algebraically, it takes you from one simplex tableau to another in which the lower right corner either increases in the case of maximization or decreases in the case of minimization.

I will continue writing the simplex tableau in such a way that the simple columns having only one entry nonzero are on the left. As explained above, this amounts to permuting the variables. I will do this because it is possible to describe what is going on without onerous notation. However, in the examples, I won't worry so much about it. Thus, from a basic feasible solution, a simplex tableau of the following form has been obtained in which the columns for the basic variables, $\mathbf{x}_B$ are listed first and $\mathbf{b} \geq \mathbf{0}$.

$$\begin{pmatrix} I & F & \mathbf{0} & \mathbf{b} \\ \mathbf{0} & \mathbf{c} & 1 & z^0 \end{pmatrix} \tag{9.10}$$

Let $x_i^0 = b_i$ for $i = 1, \cdots, m$ and $x_i^0 = 0$ for $i > m$. Then $\left(\mathbf{x}^0, z^0\right)$ is a solution to the above system and since $\mathbf{b} \geq \mathbf{0}$, it follows $\left(\mathbf{x}^0, z^0\right)$ is a basic feasible solution.

If $c_i < 0$ for some i, and if $F_{ji} \leq 0$ so that a whole column of $\begin{pmatrix} F \\ \mathbf{c} \end{pmatrix}$ is ≤ 0 with the bottom entry < 0, then letting x_i be the variable corresponding to that column, you could

leave all the other entries of $\mathbf{x}_F$ equal to zero but change x_i to be positive. Let the new vector be denoted by $\mathbf{x}'_F$ and letting $\mathbf{x}'_B = \mathbf{b} - F\mathbf{x}'_F$ it follows

$$\begin{aligned}(\mathbf{x}'_B)_k &= b_k - \sum_j F_{kj} (\mathbf{x}_F)_j \\ &= b_k - F_{ki} x_i \geq 0\end{aligned}$$

Now this shows $(\mathbf{x}'_B, \mathbf{x}'_F)$ is feasible whenever $x_i > 0$ and so you could let x_i become arbitrarily large and positive and conclude there is no maximum for z because

$$z = -\mathbf{c}\mathbf{x}'_F + z^0 = (-c_i) x_i + z^0 \tag{9.11}$$

If this happens in a simplex tableau, you can say there is no maximum and stop.

What if $\mathbf{c} \geq \mathbf{0}$? Then $z = z^0 - \mathbf{c}\mathbf{x}_F$ and to satisfy the constraints, $\mathbf{x}_F \geq 0$. Therefore, in this case, z^0 is the largest possible value of z and so the maximum has been found. You stop when this occurs. Next I explain what to do if neither of the above stopping conditions hold.

The only case which remains is that some $c_i < 0$ and some $F_{ji} > 0$. You pick a column in $\begin{pmatrix} F \\ \mathbf{c} \end{pmatrix}$ in which $c_i < 0$, usually the one for which c_i is the largest in absolute value. You pick $F_{ji} > 0$ as a pivot element, divide the j^{th} row by F_{ji} and then use to obtain zeros above F_{ji} and below F_{ji}, thus obtaining a new simple column. This row operation also makes exactly one of the other simple columns into a nonsimple column. (In terms of variables, it is said that a free variable becomes a basic variable and a basic variable becomes a free variable.) Now permuting the columns and variables, yields

$$\begin{pmatrix} I & F' & \mathbf{0} & \mathbf{b}' \\ \mathbf{0} & \mathbf{c}' & 1 & z^{0\prime} \end{pmatrix}$$

where $z^{0\prime} \geq z^0$ because $z^{0\prime} = z^0 - c_i \left(\frac{b_j}{F_{ji}}\right)$ and $c_i < 0$. If $\mathbf{b}' \geq 0$, you are in the same position you were at the beginning but now z^0 is larger. Now here is the **important** thing. You don't pick just any F_{ji} when you do these row operations. You **pick the positive one for which the row operation results in $\mathbf{b}' \geq \mathbf{0}$**. Otherwise the obvious basic feasible solution obtained by letting $\mathbf{x}'_F = \mathbf{0}$ will fail to satisfy the constraint that $\mathbf{x} \geq \mathbf{0}$.

How is this done? You need

$$b'_p \equiv b_p - \frac{F_{pi} b_j}{F_{ji}} \geq 0 \tag{9.12}$$

for each $p = 1, \cdots, m$ or equivalently,

$$b_p \geq \frac{F_{pi} b_j}{F_{ji}}. \tag{9.13}$$

Now if $F_{pi} \leq 0$ the above holds. Therefore, you only need to check F_{pi} for $F_{pi} > 0$. The pivot, F_{ji} is the one which makes the quotients of the form

$$\frac{b_p}{F_{pi}}$$

for all positive F_{pi} the smallest. Having gotten a new simplex tableau, you do the same thing to it which was just done and continue. As long as $\mathbf{b} > \mathbf{0}$, so you don't encounter the degenerate case, the values for z associated with setting $\mathbf{x}_F = \mathbf{0}$ keep getting strictly larger every time the process is repeated. You keep going until you find $\mathbf{c} \geq \mathbf{0}$. Then you stop. You are at a maximum. Problems can occur in the process in the so called degenerate case when at some stage of the process some $b_j = 0$. In this case you can cycle through different values for $\mathbf{x}$ with no improvement in z. This case will not be discussed here.

9.3.2 Minimums

How does it differ if you are finding a minimum? From a basic feasible solution, a simplex tableau of the following form has been obtained in which the simple columns for the basic variables, $\mathbf{x}_B$ are listed first and $\mathbf{b} \geq \mathbf{0}$.

$$\begin{pmatrix} I & F & \mathbf{0} & \mathbf{b} \\ \mathbf{0} & \mathbf{c} & 1 & z^0 \end{pmatrix} \tag{9.14}$$

Let $x_i^0 = b_i$ for $i = 1, \cdots, m$ and $x_i^0 = 0$ for $i > m$. Then $\left(\mathbf{x}^0, z^0\right)$ is a solution to the above system and since $\mathbf{b} \geq \mathbf{0}$, it follows $\left(\mathbf{x}^0, z^0\right)$ is a basic feasible solution. So far, there is no change.

Suppose first that some $c_i > 0$ and $F_{ji} \leq 0$ for each j. Then let $\mathbf{x}_F'$ consist of changing x_i by making it positive but leaving the other entries of $\mathbf{x}_F$ equal to 0. Then from the bottom row,

$$z = -\mathbf{c}\mathbf{x}_F' + z^0 = -c_i x_i + z^0$$

and you let $\mathbf{x}_B' = \mathbf{b} - F\mathbf{x}_F' \geq \mathbf{0}$. Thus the constraints continue to hold when x_i is made increasingly positive and it follows from the above equation that there is no minimum for z. You stop when this happens.

Next suppose $\mathbf{c} \leq \mathbf{0}$. Then in this case, $z = z^0 - \mathbf{c}\mathbf{x}_F$ and from the constraints, $\mathbf{x}_F \geq \mathbf{0}$ and so $-\mathbf{c}\mathbf{x}_F \geq 0$ and so z^0 is the minimum value and you stop since this is what you are looking for.

What do you do in the case where some $c_i > 0$ and some $F_{ji} > 0$? In this case, you use the simplex algorithm as in the case of maximums to obtain a new simplex tableau in which $z^{0\prime}$ is smaller. You choose F_{ji} the same way to be the positive entry of the i^{th} column such that $b_p/F_{pi} \geq b_j/F_{ji}$ for all positive entries, F_{pi} and do the same row operations. Now this time,

$$z^{0\prime} = z^0 - c_i \left(\frac{b_j}{F_{ji}}\right) < z^0$$

As in the case of maximums no problem can occur and the process will converge unless you have the degenerate case in which some $b_j = 0$. As in the earlier case, this is most unfortunate when it occurs. You see what happens of course. z^0 does not change and the algorithm just delivers different values of the variables forever with no improvement.

To summarize the geometrical significance of the simplex algorithm, it takes you from one corner of the feasible region to another. You go in one direction to find the maximum and in another to find the minimum. For the maximum you try to get rid of negative entries of $\mathbf{c}$ and for minimums you try to eliminate positive entries of $\mathbf{c}$ where the method of elimination involves the auspicious use of an appropriate pivot element and row operations.

Now return to Example 9.2.2. It will be modified to be a maximization problem.

Example 9.3.1 *Maximize $z = x_1 - x_2$ subject to the constraints, $x_1 + 2x_2 \leq 10, x_1 + 2x_2 \geq 2$, and $2x_1 + x_2 \leq 6, x_i \geq 0$.*

Recall this is the same as maximizing $z = x_1 - x_2$ subject to

$$\begin{pmatrix} 1 & 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & -1 & 0 \\ 2 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 10 \\ 2 \\ 6 \end{pmatrix}, \mathbf{x} \geq \mathbf{0},$$

the variables, x_3, x_4, x_5 being slack variables. Recall the simplex tableau was

$$\begin{pmatrix} 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 5 \\ 0 & 1 & 0 & 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & \frac{3}{2} & -\frac{1}{2} & 0 & 1 \\ 0 & 0 & 0 & -\frac{3}{2} & -\frac{1}{2} & 1 & -5 \end{pmatrix}$$

with the variables ordered as x_2, x_4, x_5, x_1, x_3 and so $\mathbf{x}_B = (x_2, x_4, x_5)$ and $\mathbf{x}_F = (x_1, x_3)$.

Apply the simplex algorithm to the fourth column because $-\frac{3}{2} < 0$ and this is the most negative entry in the bottom row. The pivot is $3/2$ because $1/(3/2) = 2/3 < 5/(1/2)$. Dividing this row by $3/2$ and then using this to zero out the other elements in that column, the new simplex tableau is

$$\begin{pmatrix} 1 & 0 & -\frac{1}{3} & 0 & \frac{2}{3} & 0 & \frac{14}{3} \\ 0 & 1 & 0 & 0 & 1 & 0 & 8 \\ 0 & 0 & \frac{2}{3} & 1 & -\frac{1}{3} & 0 & \frac{2}{3} \\ 0 & 0 & 1 & 0 & -1 & 1 & -4 \end{pmatrix}.$$

Now there is still a negative number in the bottom left row. Therefore, the process should be continued. This time the pivot is the $2/3$ in the top of the column. Dividing the top row by $2/3$ and then using this to zero out the entries below it,

$$\begin{pmatrix} \frac{3}{2} & 0 & -\frac{1}{2} & 0 & 1 & 0 & 7 \\ -\frac{3}{2} & 1 & \frac{1}{2} & 0 & 0 & 0 & 1 \\ \frac{1}{2} & 0 & \frac{1}{2} & 1 & 0 & 0 & 3 \\ \frac{3}{2} & 0 & \frac{1}{2} & 0 & 0 & 1 & 3 \end{pmatrix}.$$

Now all the numbers on the bottom left row are nonnegative so the process stops. Now recall the variables and columns were ordered as x_2, x_4, x_5, x_1, x_3. The solution in terms of x_1 and x_2 is $x_2 = 0$ and $x_1 = 3$ and $z = 3$. Note that in the above, I did not worry about permuting the columns to keep those which go with the basic variables on the left.

Here is a bucolic example.

Example 9.3.2 *Consider the following table.*

	F_1	F_2	F_3	F_4
iron	*1*	*2*	*1*	*3*
protein	*5*	*3*	*2*	*1*
folic acid	*1*	*2*	*2*	*1*
copper	*2*	*1*	*1*	*1*
calcium	*1*	*1*	*1*	*1*

This information is available to a pig farmer and F_i denotes a particular feed. The numbers in the table contain the number of units of a particular nutrient contained in one pound of the given feed. Thus F_2 has 2 units of iron in one pound. Now suppose the cost of each feed in cents per pound is given in the following table.

F_1	F_2	F_3	F_4
2	*3*	*2*	*3*

A typical pig needs 5 units of iron, 8 of protein, 6 of folic acid, 7 of copper and 4 of calcium. (The units may change from nutrient to nutrient.) How many pounds of each feed per pig should the pig farmer use in order to minimize his cost?

His problem is to minimize $C \equiv 2x_1 + 3x_2 + 2x_3 + 3x_4$ subject to the constraints

$$\begin{aligned} x_1 + 2x_2 + x_3 + 3x_4 &\geq 5, \\ 5x_1 + 3x_2 + 2x_3 + x_4 &\geq 8, \\ x_1 + 2x_2 + 2x_3 + x_4 &\geq 6, \\ 2x_1 + x_2 + x_3 + x_4 &\geq 7, \\ x_1 + x_2 + x_3 + x_4 &\geq 4. \end{aligned}$$

where each $x_i \geq 0$. Add in the slack variables,

$$\begin{aligned} x_1 + 2x_2 + x_3 + 3x_4 - x_5 &= 5 \\ 5x_1 + 3x_2 + 2x_3 + x_4 - x_6 &= 8 \\ x_1 + 2x_2 + 2x_3 + x_4 - x_7 &= 6 \\ 2x_1 + x_2 + x_3 + x_4 - x_8 &= 7 \\ x_1 + x_2 + x_3 + x_4 - x_9 &= 4 \end{aligned}$$

The augmented matrix for this system is

$$\begin{pmatrix} 1 & 2 & 1 & 3 & -1 & 0 & 0 & 0 & 0 & 5 \\ 5 & 3 & 2 & 1 & 0 & -1 & 0 & 0 & 0 & 8 \\ 1 & 2 & 2 & 1 & 0 & 0 & -1 & 0 & 0 & 6 \\ 2 & 1 & 1 & 1 & 0 & 0 & 0 & -1 & 0 & 7 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & 4 \end{pmatrix}$$

How in the world can you find a basic feasible solution? Remember the simplex algorithm is designed to keep the entries in the right column nonnegative so you use this algorithm a few times till the obvious solution is a basic feasible solution.

Consider the first column. The pivot is the 5. Using the row operations described in the algorithm, you get

$$\begin{pmatrix} 0 & \frac{7}{5} & \frac{3}{5} & \frac{14}{5} & -1 & \frac{1}{5} & 0 & 0 & 0 & \frac{17}{5} \\ 1 & \frac{3}{5} & \frac{2}{5} & \frac{1}{5} & 0 & -\frac{1}{5} & 0 & 0 & 0 & \frac{8}{5} \\ 0 & \frac{7}{5} & \frac{8}{5} & \frac{4}{5} & 0 & \frac{1}{5} & -1 & 0 & 0 & \frac{22}{5} \\ 0 & -\frac{1}{5} & \frac{1}{5} & \frac{3}{5} & 0 & \frac{2}{5} & 0 & -1 & 0 & \frac{19}{5} \\ 0 & \frac{2}{5} & \frac{3}{5} & \frac{4}{5} & 0 & \frac{1}{5} & 0 & 0 & -1 & \frac{12}{5} \end{pmatrix}$$

Now go to the second column. The pivot in this column is the 7/5. This is in a different row than the pivot in the first column so I will use it to zero out everything below it. This will get rid of the zeros in the fifth column and introduce zeros in the second. This yields

$$\begin{pmatrix} 0 & 1 & \frac{3}{7} & 2 & -\frac{5}{7} & \frac{1}{7} & 0 & 0 & 0 & \frac{17}{7} \\ 1 & 0 & \frac{1}{7} & -1 & \frac{3}{7} & -\frac{2}{7} & 0 & 0 & 0 & \frac{1}{7} \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & \frac{2}{7} & 1 & -\frac{1}{7} & \frac{3}{7} & 0 & -1 & 0 & \frac{30}{7} \\ 0 & 0 & \frac{3}{7} & 0 & \frac{2}{7} & \frac{1}{7} & 0 & 0 & -1 & \frac{10}{7} \end{pmatrix}$$

Now consider another column, this time the fourth. I will pick this one because it has some negative numbers in it so there are fewer entries to check in looking for a pivot. Unfortunately, the pivot is the top 2 and I don't want to pivot on this because it would destroy the zeros in the second column. Consider the fifth column. It is also not a good choice because the pivot is the second element from the top and this would destroy the zeros

in the first column. Consider the sixth column. I can use either of the two bottom entries as the pivot. The matrix is

$$\begin{pmatrix} 0 & 1 & 0 & 2 & -1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & -1 & 1 & 0 & 0 & 0 & -2 & 3 \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 & -1 & 0 & 0 & -1 & 3 & 0 \\ 0 & 0 & 3 & 0 & 2 & 1 & 0 & 0 & -7 & 10 \end{pmatrix}$$

Next consider the third column. The pivot is the 1 in the third row. This yields

$$\begin{pmatrix} 0 & 1 & 0 & 2 & -1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & -2 & 2 \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 & 0 & 0 & -1 & -1 & 3 & 1 \\ 0 & 0 & 0 & 6 & -1 & 1 & 3 & 0 & -7 & 7 \end{pmatrix}.$$

There are still 5 columns which consist entirely of zeros except for one entry. Four of them have that entry equal to 1 but one still has a -1 in it, the -1 being in the fourth column. I need to do the row operations on a nonsimple column which has the pivot in the fourth row. Such a column is the second to the last. The pivot is the 3. The new matrix is

$$\begin{pmatrix} 0 & 1 & 0 & \frac{7}{3} & -1 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{2}{3} \\ 1 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & -\frac{2}{3} & 0 & \frac{8}{3} \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\frac{1}{3} & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & 1 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{11}{3} & -1 & 1 & \frac{2}{3} & -\frac{7}{3} & 0 & \frac{28}{3} \end{pmatrix}. \tag{9.15}$$

Now the obvious basic solution is feasible. You let $x_4 = 0 = x_5 = x_7 = x_8$ and $x_1 = 8/3, x_2 = 2/3, x_3 = 1$, and $x_6 = 28/3$. You don't need to worry too much about this. It is the above matrix which is desired. Now you can assemble the simplex tableau and begin the algorithm. Remember $C \equiv 2x_1 + 3x_2 + 2x_3 + 3x_4$. First add the row and column which deal with C. This yields

$$\begin{pmatrix} 0 & 1 & 0 & \frac{7}{3} & -1 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 1 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & -\frac{2}{3} & 0 & 0 & \frac{8}{3} \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\frac{1}{3} & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{11}{3} & -1 & 1 & \frac{2}{3} & -\frac{7}{3} & 0 & 0 & \frac{28}{3} \\ -2 & -3 & -2 & -3 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{9.16}$$

Now you do row operations to keep the simple columns of 9.15 simple in 9.16. Of course you could permute the columns if you wanted but this is not necessary.

This yields the following for a simplex tableau. Now it is a matter of getting rid of the positive entries in the bottom row because you are trying to minimize.

$$\begin{pmatrix} 0 & 1 & 0 & \frac{7}{3} & -1 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 1 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & -\frac{2}{3} & 0 & 0 & \frac{8}{3} \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\frac{1}{3} & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{11}{3} & -1 & 1 & \frac{2}{3} & -\frac{7}{3} & 0 & 0 & \frac{28}{3} \\ 0 & 0 & 0 & \frac{2}{3} & -1 & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & 1 & \frac{28}{3} \end{pmatrix}$$

The most positive of them is the 2/3 and so I will apply the algorithm to this one first. The pivot is the 7/3. After doing the row operation the next tableau is

$$\begin{pmatrix} 0 & \frac{3}{7} & 0 & 1 & -\frac{3}{7} & 0 & \frac{1}{7} & \frac{1}{7} & 0 & 0 & \frac{2}{7} \\ 1 & -\frac{1}{7} & 0 & 0 & \frac{1}{7} & 0 & \frac{2}{7} & -\frac{5}{7} & 0 & 0 & \frac{18}{7} \\ 0 & \frac{6}{7} & 1 & 0 & \frac{1}{7} & 0 & -\frac{5}{7} & \frac{2}{7} & 0 & 0 & \frac{11}{7} \\ 0 & \frac{1}{7} & 0 & 0 & -\frac{1}{7} & 0 & -\frac{2}{7} & -\frac{2}{7} & 1 & 0 & \frac{3}{7} \\ 0 & -\frac{11}{7} & 0 & 0 & \frac{4}{7} & 1 & \frac{1}{7} & -\frac{20}{7} & 0 & 0 & \frac{58}{7} \\ 0 & -\frac{2}{7} & 0 & 0 & -\frac{5}{7} & 0 & -\frac{3}{7} & -\frac{3}{7} & 0 & 1 & \frac{64}{7} \end{pmatrix}$$

and you see that all the entries are negative and so the minimum is 64/7 and it occurs when $x_1 = 18/7, x_2 = 0, x_3 = 11/7, x_4 = 2/7$.

There is no maximum for the above problem. However, I will pretend I don't know this and attempt to use the simplex algorithm. You set up the simiplex tableau the same way. Recall it is

$$\begin{pmatrix} 0 & 1 & 0 & \frac{7}{3} & -1 & 0 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 1 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & -\frac{2}{3} & 0 & 0 & \frac{8}{3} \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\frac{1}{3} & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{11}{3} & -1 & 1 & \frac{2}{3} & -\frac{7}{3} & 0 & 0 & \frac{28}{3} \\ 0 & 0 & 0 & \frac{2}{3} & -1 & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & 1 & \frac{28}{3} \end{pmatrix}$$

Now to maximize, you try to get rid of the negative entries in the bottom left row. The most negative entry is the -1 in the fifth column. The pivot is the 1 in the third row of this column. The new tableau is

$$\begin{pmatrix} 0 & 1 & 1 & \frac{1}{3} & 0 & 0 & -\frac{2}{3} & \frac{1}{3} & 0 & 0 & \frac{5}{3} \\ 1 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} & -\frac{2}{3} & 0 & 0 & \frac{8}{3} \\ 0 & 0 & 1 & -2 & 1 & 0 & -1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\frac{1}{3} & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & 1 & 0 & \frac{1}{3} \\ 0 & 0 & 1 & \frac{5}{3} & 0 & 1 & -\frac{1}{3} & -\frac{7}{3} & 0 & 0 & \frac{31}{3} \\ 0 & 0 & 1 & -\frac{4}{3} & 0 & 0 & -\frac{4}{3} & -\frac{1}{3} & 0 & 1 & \frac{31}{3} \end{pmatrix}.$$

Consider the fourth column. The pivot is the top 1/3. The new tableau is

$$\begin{pmatrix} 0 & 3 & 3 & 1 & 0 & 0 & -2 & 1 & 0 & 0 & 5 \\ 1 & -1 & -1 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 1 \\ 0 & 6 & 7 & 0 & 1 & 0 & -5 & 2 & 0 & 0 & 11 \\ 0 & 1 & 1 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 2 \\ 0 & -5 & -4 & 0 & 0 & 1 & 3 & -4 & 0 & 0 & 2 \\ 0 & 4 & 5 & 0 & 0 & 0 & -4 & 1 & 0 & 1 & 17 \end{pmatrix}$$

There is still a negative in the bottom, the -4. The pivot in that column is the 3. The algorithm yields

$$\begin{pmatrix} 0 & -\frac{1}{3} & \frac{1}{3} & 1 & 0 & \frac{2}{3} & 0 & -\frac{5}{3} & 0 & 0 & \frac{19}{3} \\ 1 & \frac{2}{3} & \frac{1}{3} & 0 & 0 & -\frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ 0 & -\frac{7}{3} & \frac{1}{3} & 0 & 1 & \frac{5}{3} & 0 & -\frac{14}{3} & 0 & 0 & \frac{43}{3} \\ 0 & -\frac{2}{3} & -\frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 & -\frac{4}{3} & 1 & 0 & \frac{8}{3} \\ 0 & -\frac{5}{3} & -\frac{4}{3} & 0 & 0 & \frac{1}{3} & 1 & -\frac{4}{3} & 0 & 0 & \frac{2}{3} \\ 0 & -\frac{8}{3} & -\frac{1}{3} & 0 & 0 & \frac{4}{3} & 0 & -\frac{13}{3} & 0 & 1 & \frac{59}{3} \end{pmatrix}$$

Note how z keeps getting larger. Consider the column having the $-13/3$ in it. The pivot is

the single positive entry, 1/3. The next tableau is

$$\begin{pmatrix} 5 & 3 & 2 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 8 \\ 3 & 2 & 1 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 1 \\ 14 & 7 & 5 & 0 & 1 & -3 & 0 & 0 & 0 & 0 & 19 \\ 4 & 2 & 1 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 4 \\ 4 & 1 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 2 \\ 13 & 6 & 4 & 0 & 0 & -3 & 0 & 0 & 0 & 1 & 24 \end{pmatrix}.$$

There is a column consisting of all negative entries. There is therefore, no maximum. Note also how there is no way to pick the pivot in that column.

Example 9.3.3 *Minimize* $z = x_1 - 3x_2 + x_3$ *subject to the constraints* $x_1 + x_2 + x_3 \leq 10, x_1 + x_2 + x_3 \geq 2$, $x_1 + x_2 + 3x_3 \leq 8$ *and* $x_1 + 2x_2 + x_3 \leq 7$ *with all variables nonnegative.*

There exists an answer because the region defined by the constraints is closed and bounded. Adding in slack variables you get the following augmented matrix corresponding to the constraints.

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 10 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 2 \\ 1 & 1 & 3 & 0 & 0 & 1 & 0 & 8 \\ 1 & 2 & 1 & 0 & 0 & 0 & 1 & 7 \end{pmatrix}$$

Of course there is a problem with the obvious solution obtained by setting to zero all variables corresponding to a nonsimple column because of the simple column which has the -1 in it. Therefore, I will use the simplex algorithm to make this column non simple. The third column has the 1 in the second row as the pivot so I will use this column. This yields

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 8 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 2 \\ -2 & -2 & 0 & 0 & 3 & 1 & 0 & 2 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 5 \end{pmatrix} \tag{9.17}$$

and the obvious solution is feasible. Now it is time to assemble the simplex tableau. First add in the bottom row and second to last column corresponding to the equation for z. This yields

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 8 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & 2 \\ -2 & -2 & 0 & 0 & 3 & 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 5 \\ -1 & 3 & -1 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Next you need to zero out the entries in the bottom row which are below one of the simple columns in 9.17. This yields the simplex tableau

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 8 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & 2 \\ -2 & -2 & 0 & 0 & 3 & 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 5 \\ 0 & 4 & 0 & 0 & -1 & 0 & 0 & 1 & 2 \end{pmatrix}.$$

The desire is to minimize this so you need to get rid of the positive entries in the left bottom row. There is only one such entry, the 4. In that column the pivot is the 1 in the second

row of this column. Thus the next tableau is

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 8 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & 2 \\ 0 & 0 & 2 & 0 & 1 & 1 & 0 & 0 & 6 \\ -1 & 0 & -1 & 0 & 2 & 0 & 1 & 0 & 3 \\ -4 & 0 & -4 & 0 & 3 & 0 & 0 & 1 & -6 \end{pmatrix}$$

There is still a positive number there, the 3. The pivot in this column is the 2. Apply the algorithm again. This yields

$$\begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 1 & 0 & 0 & -\frac{1}{2} & 0 & \frac{13}{2} \\ \frac{1}{2} & 1 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{7}{2} \\ \frac{1}{2} & 0 & \frac{5}{2} & 0 & 0 & 1 & -\frac{1}{2} & 0 & \frac{3}{2} \\ -\frac{1}{2} & 0 & -\frac{1}{2} & 0 & 1 & 0 & \frac{1}{2} & 0 & \frac{3}{2} \\ -\frac{5}{2} & 0 & -\frac{5}{2} & 0 & 0 & 0 & -\frac{3}{2} & 1 & -\frac{21}{2} \end{pmatrix}.$$

Now all the entries in the left bottom row are nonpositive so the process has stopped. The minimum is $-21/2$. It occurs when $x_1 = 0$, $x_2 = 7/2, x_3 = 0$.

Now consider the same problem but change the word, minimize to the word, maximize.

Example 9.3.4 *Maximize $z = x_1 - 3x_2 + x_3$ subject to the constraints $x_1 + x_2 + x_3 \leq 10, x_1 + x_2 + x_3 \geq 2$, $x_1 + x_2 + 3x_3 \leq 8$ and $x_1 + 2x_2 + x_3 \leq 7$ with all variables nonnegative.*

The first part of it is the same. You wind up with the same simplex tableau,

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 8 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 0 & 2 \\ -2 & -2 & 0 & 0 & 3 & 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 5 \\ 0 & 4 & 0 & 0 & -1 & 0 & 0 & 1 & 2 \end{pmatrix}$$

but this time, you apply the algorithm to get rid of the negative entries in the left bottom row. There is a -1. Use this column. The pivot is the 3. The next tableau is

$$\begin{pmatrix} \frac{2}{3} & \frac{2}{3} & 0 & 1 & 0 & -\frac{1}{3} & 0 & 0 & \frac{22}{3} \\ \frac{1}{3} & \frac{1}{3} & 1 & 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{8}{3} \\ -\frac{2}{3} & -\frac{2}{3} & 0 & 0 & 1 & \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ \frac{2}{3} & \frac{5}{3} & 0 & 0 & 0 & -\frac{1}{3} & 1 & 0 & \frac{13}{3} \\ -\frac{2}{3} & \frac{10}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 1 & \frac{8}{3} \end{pmatrix}$$

There is still a negative entry, the $-2/3$. This will be the new pivot column. The pivot is the $2/3$ on the fourth row. This yields

$$\begin{pmatrix} 0 & -1 & 0 & 1 & 0 & 0 & -1 & 0 & 3 \\ 0 & -\frac{1}{2} & 1 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 5 \\ 1 & \frac{5}{2} & 0 & 0 & 0 & -\frac{1}{2} & \frac{3}{2} & 0 & \frac{13}{2} \\ 0 & 5 & 0 & 0 & 0 & 0 & 1 & 1 & 7 \end{pmatrix}$$

and the process stops. The maximum for z is 7 and it occurs when $x_1 = 13/2, x_2 = 0, x_3 = 1/2$.

9.4 Finding A Basic Feasible Solution

By now it should be fairly clear that finding a basic feasible solution can create considerable difficulty. Indeed, given a system of linear inequalities along with the requirement that each variable be nonnegative, do there even exist points satisfying all these inequalities? If you have many variables, you can't answer this by drawing a picture. Is there some other way to do this which is more systematic than what was presented above? The answer is yes. It is called the method of artificial variables. I will illustrate this method with an example.

Example 9.4.1 *Find a basic feasible solution to the system* $2x_1+x_2-x_3 \geq 3, x_1+x_2+x_3 \geq 2, x_1+x_2+x_3 \leq 7$ *and* $\mathbf{x} \geq \mathbf{0}$.

If you write the appropriate augmented matrix with the slack variables,

$$\begin{pmatrix} 2 & 1 & -1 & -1 & 0 & 0 & 3 \\ 1 & 1 & 1 & 0 & -1 & 0 & 2 \\ 1 & 1 & 1 & 0 & 0 & 1 & 7 \end{pmatrix} \tag{9.18}$$

The obvious solution is not feasible. This is why it would be hard to get started with the simplex method. What is the problem? It is those -1 entries in the fourth and fifth columns. To get around this, you add in artificial variables to get an augmented matrix of the form

$$\begin{pmatrix} 2 & 1 & -1 & -1 & 0 & 0 & 1 & 0 & 3 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 1 & 2 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 7 \end{pmatrix} \tag{9.19}$$

Thus the variables are $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$. Suppose you can find a feasible solution to the system of equations represented by the above augmented matrix. Thus all variables are nonnegative. Suppose also that it can be done in such a way that x_8 and x_7 happen to be 0. Then it will follow that $x_1, \cdots, x_6$ is a feasible solution for 9.18. Conversely, if you can find a feasible solution for 9.18, then letting x_7 and x_8 both equal zero, you have obtained a feasible solution to 9.19. Since all variables are nonnegative, x_7 and x_8 both equalling zero is equivalent to saying the minimum of $z = x_7 + x_8$ subject to the constraints represented by the above augmented matrix equals zero. This has proved the following simple observation.

Observation 9.4.2 *There exists a feasible solution to the constraints represented by the augmented matrix of 9.18 and* $\mathbf{x} \geq \mathbf{0}$ *if and only if the minimum of* $x_7 + x_8$ *subject to the constraints of 9.19 and* $\mathbf{x} \geq \mathbf{0}$ *exists and equals* 0.

Of course a similar observation would hold in other similar situations. Now the point of all this is that it is trivial to see a feasible solution to 9.19, namely $x_6 = 7, x_7 = 3, x_8 = 2$ and all the other variables may be set to equal zero. Therefore, it is easy to find an initial simplex tableau for the minimization problem just described. First add the column and row for z

$$\begin{pmatrix} 2 & 1 & -1 & -1 & 0 & 0 & 1 & 0 & 0 & 3 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 1 & 0 & 2 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 7 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 1 & 0 \end{pmatrix}$$

Next it is necessary to make the last two columns on the bottom left row into simple columns. Performing the row operation, this yields an initial simplex tableau,

$$\begin{pmatrix} 2 & 1 & -1 & -1 & 0 & 0 & 1 & 0 & 0 & 3 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 & 1 & 0 & 2 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 7 \\ 3 & 2 & 0 & -1 & -1 & 0 & 0 & 0 & 1 & 5 \end{pmatrix}$$

Now the algorithm involves getting rid of the positive entries on the left bottom row. Begin with the first column. The pivot is the 2. An application of the simplex algorithm yields the new tableau

$$\begin{pmatrix} 1 & \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 & \frac{3}{2} \\ 0 & \frac{1}{2} & \frac{3}{2} & \frac{1}{2} & -1 & 0 & -\frac{1}{2} & 1 & 0 & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{3}{2} & \frac{1}{2} & 0 & 1 & -\frac{1}{2} & 0 & 0 & \frac{11}{2} \\ 0 & \frac{1}{2} & \frac{3}{2} & \frac{1}{2} & -1 & 0 & -\frac{3}{2} & 0 & 1 & \frac{1}{2} \end{pmatrix}$$

Now go to the third column. The pivot is the $3/2$ in the second row. An application of the simplex algorithm yields

$$\begin{pmatrix} 1 & \frac{2}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{5}{3} \\ 0 & \frac{1}{3} & 1 & \frac{1}{3} & -\frac{2}{3} & 0 & -\frac{1}{3} & \frac{2}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & -1 & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 1 & 0 \end{pmatrix} \tag{9.20}$$

and you see there are only nonpositive numbers on the bottom left column so the process stops and yields 0 for the minimum of $z = x_7 + x_8$. As for the other variables, $x_1 = 5/3, x_2 = 0, x_3 = 1/3, x_4 = 0, x_5 = 0, x_6 = 5$. Now as explained in the above observation, this is a basic feasible solution for the original system 9.18.

Now consider a maximization problem associated with the above constraints.

Example 9.4.3 *Maximize* $x_1 - x_2 + 2x_3$ *subject to the constraints,* $2x_1 + x_2 - x_3 \geq 3, x_1 + x_2 + x_3 \geq 2, x_1 + x_2 + x_3 \leq 7$ *and* $\mathbf{x} \geq \mathbf{0}$.

From 9.20 you can immediately assemble an initial simplex tableau. You begin with the first 6 columns and top 3 rows in 9.20. Then add in the column and row for z. This yields

$$\begin{pmatrix} 1 & \frac{2}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & \frac{5}{3} \\ 0 & \frac{1}{3} & 1 & \frac{1}{3} & -\frac{2}{3} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 5 \\ -1 & 1 & -2 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

and you first do row operations to make the first and third columns simple columns. Thus the next simplex tableau is

$$\begin{pmatrix} 1 & \frac{2}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & \frac{5}{3} \\ 0 & \frac{1}{3} & 1 & \frac{1}{3} & -\frac{2}{3} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 5 \\ 0 & \frac{7}{3} & 0 & \frac{1}{3} & -\frac{5}{3} & 0 & 1 & \frac{7}{3} \end{pmatrix}$$

You are trying to get rid of negative entries in the bottom left row. There is only one, the $-5/3$. The pivot is the 1. The next simplex tableau is then

$$\begin{pmatrix} 1 & \frac{2}{3} & 0 & -\frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{10}{3} \\ 0 & \frac{1}{3} & 1 & \frac{1}{3} & 0 & \frac{2}{3} & 0 & \frac{11}{3} \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 5 \\ 0 & \frac{7}{3} & 0 & \frac{1}{3} & 0 & \frac{5}{3} & 1 & \frac{32}{3} \end{pmatrix}$$

and so the maximum value of z is $32/3$ and it occurs when $x_1 = 10/3, x_2 = 0$ and $x_3 = 11/3$.

9.5 Duality

You can solve minimization problems by solving maximization problems. You can also go the other direction and solve maximization problems by minimization problems. Sometimes this makes things much easier. To be more specific, the two problems to be considered are

A.) Minimize $z = \mathbf{cx}$ subject to $\mathbf{x} \geq \mathbf{0}$ and $A\mathbf{x} \geq \mathbf{b}$ and
B.) Maximize $w = \mathbf{yb}$ such that $\mathbf{y} \geq \mathbf{0}$ and $\mathbf{y}A \leq \mathbf{c}$,

$$\left(\text{equivalently } A^T\mathbf{y}^T \geq \mathbf{c}^T \text{ and } w = \mathbf{b}^T\mathbf{y}^T\right).$$

In these problems it is assumed A is an $m \times p$ matrix.

I will show how a solution of the first yields a solution of the second and then show how a solution of the second yields a solution of the first. The problems, A.) and B.) are called dual problems.

Lemma 9.5.1 *Let* $\mathbf{x}$ *be a solution of the inequalities of* A.) *and let* $\mathbf{y}$ *be a solution of the inequalities of* B.). *Then*

$$\mathbf{cx} \geq \mathbf{yb}.$$

and if equality holds in the above, then $\mathbf{x}$ *is the solution to* A.) *and* $\mathbf{y}$ *is a solution to* B.).

Proof: This follows immediately. Since $\mathbf{c} \geq \mathbf{y}A$,

$$\mathbf{cx} \geq \mathbf{y}A\mathbf{x} \geq \mathbf{yb}.$$

It follows from this lemma that if $\mathbf{y}$ satisfies the inequalities of B.) and $\mathbf{x}$ satisfies the inequalities of A.) then if equality holds in the above lemma, it must be that $\mathbf{x}$ is a solution of A.) and $\mathbf{y}$ is a solution of B.). This proves the lemma.

Now recall that to solve either of these problems using the simplex method, you first add in slack variables. Denote by $\mathbf{x}'$ and $\mathbf{y}'$ the enlarged list of variables. Thus $\mathbf{x}'$ has at least m entries and so does $\mathbf{y}'$ and the inequalities involving A were replaced by equalities whose augmented matrices were of the form

$$\begin{pmatrix} A & -I & \mathbf{b} \end{pmatrix}, \text{ and } \begin{pmatrix} A^T & I & \mathbf{c}^T \end{pmatrix}$$

Then you included the row and column for z and w to obtain

$$\begin{pmatrix} A & -I & \mathbf{0} & \mathbf{b} \\ -\mathbf{c} & \mathbf{0} & 1 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} A^T & I & \mathbf{0} & \mathbf{c}^T \\ -\mathbf{b}^T & \mathbf{0} & 1 & 0 \end{pmatrix}. \tag{9.21}$$

Then the problems have basic feasible solutions if it is possible to permute the first $p+m$ columns in the above two matrices and obtain matrices of the form

$$\begin{pmatrix} B & F & \mathbf{0} & \mathbf{b} \\ -\mathbf{c}_B & -\mathbf{c}_F & 1 & 0 \end{pmatrix} \text{ and } \begin{pmatrix} B_1 & F_1 & \mathbf{0} & \mathbf{c}^T \\ -\mathbf{b}_{B_1}^T & -\mathbf{b}_{F_1}^T & 1 & 0 \end{pmatrix} \tag{9.22}$$

where B, B_1 are invertible $m \times m$ and $p \times p$ matrices and denoting the variables associated with these columns by $\mathbf{x}_B, \mathbf{y}_B$ and those variables associated with F or F_1 by $\mathbf{x}_F$ and $\mathbf{y}_F$, it follows that letting $B\mathbf{x}_B = \mathbf{b}$ and $\mathbf{x}_F = \mathbf{0}$, the resulting vector, $\mathbf{x}'$ is a solution to $\mathbf{x}' \geq \mathbf{0}$ and $\begin{pmatrix} A & -I \end{pmatrix}\mathbf{x}' = \mathbf{b}$ with similar constraints holding for $\mathbf{y}'$. In other words, it is possible to obtain simplex tableaus,

$$\begin{pmatrix} I & B^{-1}F & \mathbf{0} & B^{-1}\mathbf{b} \\ \mathbf{0} & \mathbf{c}_B B^{-1}F - \mathbf{c}_F & 1 & \mathbf{c}_B B^{-1}\mathbf{b} \end{pmatrix}, \begin{pmatrix} I & B_1^{-1}F_1 & \mathbf{0} & B_1^{-1}\mathbf{c}^T \\ \mathbf{0} & \mathbf{b}_{B_1}^T B_1^{-1}F - \mathbf{b}_{F_1}^T & 1 & \mathbf{b}_{B_1}^T B_1^{-1}\mathbf{c}^T \end{pmatrix} \tag{9.23}$$

Similar considerations apply to the second problem. Thus as just described, a basic feasible solution is one which determines a simplex tableau like the above in which you get a feasible solution by setting all but the first m variables equal to zero. The simplex algorithm takes you from one basic feasible solution to another till eventually, if there is no degeneracy, you obtain a basic feasible solution which yields the solution of the problem of interest.

Theorem 9.5.2 *Suppose there exists a solution,* $\mathbf{x}$ *to* $A.$*) where* $\mathbf{x}$ *is a basic feasible solution of the inequalities of* $\mathbf{A}.$*). Then there exists a solution,* $\mathbf{y}$ *to* $B.$*) and* $\mathbf{cx} = \mathbf{by}$*. It is also possible to find* $\mathbf{y}$ *from* $\mathbf{x}$ *using a simple formula.*

Proof: Since the solution to $A.$) is basic and feasible, there exists a simplex tableau like 9.23 such that $\mathbf{x}'$ can be split into $\mathbf{x}_B$ and $\mathbf{x}_F$ such that $\mathbf{x}_F = 0$ and $\mathbf{x}_B = B^{-1}\mathbf{b}$. Now since it is a minimizer, it follows $\mathbf{c}_B B^{-1}F - \mathbf{c}_F \leq \mathbf{0}$ and the minimum value for $\mathbf{cx}$ is $\mathbf{c}_B B^{-1}\mathbf{b}$. Stating this again, $\mathbf{cx} = \mathbf{c}_B B^{-1}\mathbf{b}$. Is it possible you can take $\mathbf{y} = \mathbf{c}_B B^{-1}$? From Lemma 9.5.1 this will be so if $\mathbf{c}_B B^{-1}$ solves the constraints of problem $B.$). Is $\mathbf{c}_B B^{-1} \geq 0$? Is $\mathbf{c}_B B^{-1} A \leq \mathbf{c}$? These two conditions are satisfied if and only if $\mathbf{c}_B B^{-1} \begin{pmatrix} A & -I \end{pmatrix} \leq \begin{pmatrix} \mathbf{c} & \mathbf{0} \end{pmatrix}$. Referring to the process of permuting the columns of the first augmented matrix of 9.21 to get 9.22 and doing the same permutations on the columns of $\begin{pmatrix} A & -I \end{pmatrix}$ and $\begin{pmatrix} \mathbf{c} & \mathbf{0} \end{pmatrix}$, the desired inequality holds if and only if $\mathbf{c}_B B^{-1} \begin{pmatrix} B & F \end{pmatrix} \leq \begin{pmatrix} \mathbf{c}_B & \mathbf{c}_F \end{pmatrix}$ which is equivalent to saying $\begin{pmatrix} \mathbf{c}_B & \mathbf{c}_B B^{-1}F \end{pmatrix} \leq \begin{pmatrix} \mathbf{c}_B & \mathbf{c}_F \end{pmatrix}$ and this is true because $\mathbf{c}_B B^{-1}F - \mathbf{c}_F \leq \mathbf{0}$ due to the assumption that $\mathbf{x}$ is a minimizer. The simple formula is just

$$\mathbf{y} = \mathbf{c}_B B^{-1}.$$

This proves the theorem.

The proof of the following corollary is similar.

Corollary 9.5.3 *Suppose there exists a solution,* $\mathbf{y}$ *to* $B.$*) where* $\mathbf{y}$ *is a basic feasible solution of the inequalities of* $B.$*). Then there exists a solution,* $\mathbf{x}$ *to* $A.$*) and* $\mathbf{cx} = \mathbf{by}$*. It is also possible to find* $\mathbf{x}$ *from* $\mathbf{y}$ *using a simple formula. In this case, and referring to 9.23, the simple formula is* $\mathbf{x} = B_1^{-T}\mathbf{b}_{B_1}$.

As an example, consider the pig farmers problem. The main difficulty in this problem was finding an initial simplex tableau. Now consider the following example and marvel at how all the difficulties disappear.

Example 9.5.4 *minimize* $C \equiv 2x_1 + 3x_2 + 2x_3 + 3x_4$ *subject to the constraints*

$$\begin{aligned} x_1 + 2x_2 + x_3 + 3x_4 &\geq 5, \\ 5x_1 + 3x_2 + 2x_3 + x_4 &\geq 8, \\ x_1 + 2x_2 + 2x_3 + x_4 &\geq 6, \\ 2x_1 + x_2 + x_3 + x_4 &\geq 7, \\ x_1 + x_2 + x_3 + x_4 &\geq 4. \end{aligned}$$

where each $x_i \geq 0$.

Here the dual problem is to maximize $w = 5y_1 + 8y_2 + 6y_3 + 7y_4 + 4y_5$ subject to the constraints

$$\begin{pmatrix} 1 & 5 & 1 & 2 & 1 \\ 2 & 3 & 2 & 1 & 1 \\ 1 & 2 & 2 & 1 & 1 \\ 3 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{pmatrix} \leq \begin{pmatrix} 2 \\ 3 \\ 2 \\ 3 \end{pmatrix}.$$

Adding in slack variables, these inequalities are equivalent to the system of equations whose augmented matrix is

$$\begin{pmatrix} 1 & 5 & 1 & 2 & 1 & 1 & 0 & 0 & 0 & 2 \\ 2 & 3 & 2 & 1 & 1 & 0 & 1 & 0 & 0 & 3 \\ 1 & 2 & 2 & 1 & 1 & 0 & 0 & 1 & 0 & 2 \\ 3 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 3 \end{pmatrix}$$

Now the obvious solution is feasible so there is no hunting for an initial obvious feasible solution required. Now add in the row and column for w. This yields

$$\begin{pmatrix} 1 & 5 & 1 & 2 & 1 & 1 & 0 & 0 & 0 & 0 & 2 \\ 2 & 3 & 2 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 3 \\ 1 & 2 & 2 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 2 \\ 3 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 3 \\ -5 & -8 & -6 & -7 & -4 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

It is a maximization problem so you want to eliminate the negatives in the bottom left row. Pick the column having the one which is most negative, the -8. The pivot is the top 5. Then apply the simplex algorithm to obtain

$$\begin{pmatrix} \frac{1}{5} & 1 & \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & \frac{1}{5} & 0 & 0 & 0 & 0 & \frac{2}{5} \\ \frac{7}{5} & 0 & \frac{7}{5} & -\frac{1}{5} & \frac{2}{5} & -\frac{3}{5} & 1 & 0 & 0 & 0 & \frac{9}{5} \\ \frac{3}{5} & 0 & \frac{8}{5} & \frac{1}{5} & \frac{3}{5} & -\frac{2}{5} & 0 & 1 & 0 & 0 & \frac{8}{5} \\ \frac{14}{5} & 0 & \frac{4}{5} & \frac{3}{5} & \frac{4}{5} & -\frac{1}{5} & 0 & 0 & 1 & 0 & \frac{13}{5} \\ -\frac{17}{5} & 0 & -\frac{22}{5} & -\frac{19}{5} & -\frac{12}{5} & \frac{8}{5} & 0 & 0 & 0 & 1 & \frac{16}{5} \end{pmatrix}.$$

There are still negative entries in the bottom left row. Do the simplex algorithm to the column which has the $-\frac{22}{5}$. The pivot is the $\frac{8}{5}$. This yields

$$\begin{pmatrix} \frac{1}{8} & 1 & 0 & \frac{3}{8} & \frac{1}{8} & \frac{1}{4} & 0 & -\frac{1}{8} & 0 & 0 & \frac{1}{4} \\ \frac{7}{8} & 0 & 0 & -\frac{3}{8} & -\frac{1}{8} & -\frac{1}{4} & 1 & -\frac{7}{8} & 0 & 0 & \frac{3}{4} \\ \frac{3}{8} & 0 & 1 & \frac{1}{8} & \frac{3}{8} & -\frac{1}{4} & 0 & \frac{5}{8} & 0 & 0 & \frac{3}{4} \\ \frac{5}{2} & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & -\frac{1}{2} & 1 & 0 & 2 \\ -\frac{7}{4} & 0 & 0 & -\frac{13}{4} & -\frac{3}{4} & \frac{1}{2} & 0 & \frac{11}{4} & 0 & 1 & \frac{13}{2} \end{pmatrix}$$

and there are still negative numbers. Pick the column which has the $-13/4$. The pivot is the $3/8$ in the top. This yields

$$\begin{pmatrix} \frac{1}{3} & \frac{8}{3} & 0 & 1 & \frac{1}{3} & \frac{2}{3} & 0 & -\frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 1 \\ \frac{1}{3} & -\frac{1}{3} & 1 & 0 & \frac{1}{3} & -\frac{1}{3} & 0 & \frac{2}{3} & 0 & 0 & \frac{2}{3} \\ \frac{7}{3} & -\frac{4}{3} & 0 & 0 & \frac{1}{3} & -\frac{1}{3} & 0 & -\frac{1}{3} & 1 & 0 & \frac{5}{3} \\ -\frac{2}{3} & \frac{26}{3} & 0 & 0 & \frac{1}{3} & \frac{8}{3} & 0 & \frac{5}{3} & 0 & 1 & \frac{26}{3} \end{pmatrix}$$

which has only one negative entry on the bottom left. The pivot for this first column is the $\frac{7}{3}$. The next tableau is

$$\begin{pmatrix} 0 & \frac{20}{7} & 0 & 1 & \frac{2}{7} & \frac{5}{7} & 0 & -\frac{2}{7} & -\frac{1}{7} & 0 & \frac{3}{7} \\ 0 & \frac{11}{7} & 0 & 0 & -\frac{1}{7} & \frac{1}{7} & 1 & -\frac{6}{7} & -\frac{3}{7} & 0 & \frac{2}{7} \\ 0 & -\frac{1}{7} & 1 & 0 & \frac{2}{7} & -\frac{2}{7} & 0 & \frac{5}{7} & -\frac{1}{7} & 0 & \frac{3}{7} \\ 1 & -\frac{4}{7} & 0 & 0 & \frac{1}{7} & -\frac{1}{7} & 0 & -\frac{1}{7} & \frac{3}{7} & 0 & \frac{5}{7} \\ 0 & \frac{58}{7} & 0 & 0 & \frac{3}{7} & \frac{18}{7} & 0 & \frac{11}{7} & \frac{2}{7} & 1 & \frac{64}{7} \end{pmatrix}$$

and all the entries in the left bottom row are nonnegative so the answer is $64/7$. This is the same as obtained before. So what values for $\mathbf{x}$ are needed? Here the basic variables are y_1, y_3, y_4, y_7. Consider the original augmented matrix, one step before the simplex tableau.

$$\begin{pmatrix} 1 & 5 & 1 & 2 & 1 & 1 & 0 & 0 & 0 & 0 & 2 \\ 2 & 3 & 2 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 3 \\ 1 & 2 & 2 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 2 \\ 3 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 3 \\ -5 & -8 & -6 & -7 & -4 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Permute the columns to put the columns associated with these basic variables first. Thus

$$\begin{pmatrix} 1 & 1 & 2 & 0 & 5 & 1 & 1 & 0 & 0 & 0 & 2 \\ 2 & 2 & 1 & 1 & 3 & 1 & 0 & 0 & 0 & 0 & 3 \\ 1 & 2 & 1 & 0 & 2 & 1 & 0 & 1 & 0 & 0 & 2 \\ 3 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 3 \\ -5 & -6 & -7 & 0 & -8 & -4 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

The matrix, B is

$$\begin{pmatrix} 1 & 1 & 2 & 0 \\ 2 & 2 & 1 & 1 \\ 1 & 2 & 1 & 0 \\ 3 & 1 & 1 & 0 \end{pmatrix}$$

and so B^{-T} equals

$$\begin{pmatrix} -\frac{1}{7} & -\frac{2}{7} & \frac{5}{7} & \frac{1}{7} \\ 0 & 0 & 0 & 1 \\ -\frac{1}{7} & \frac{5}{7} & -\frac{2}{7} & -\frac{6}{7} \\ \frac{3}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{3}{7} \end{pmatrix}$$

Also $\mathbf{b}_B^T = \begin{pmatrix} 5 & 6 & 7 & 0 \end{pmatrix}$ and so from Corollary 9.5.3,

$$\mathbf{x} = \begin{pmatrix} -\frac{1}{7} & -\frac{2}{7} & \frac{5}{7} & \frac{1}{7} \\ 0 & 0 & 0 & 1 \\ -\frac{1}{7} & \frac{5}{7} & -\frac{2}{7} & -\frac{6}{7} \\ \frac{3}{7} & -\frac{1}{7} & -\frac{1}{7} & -\frac{3}{7} \end{pmatrix} \begin{pmatrix} 5 \\ 6 \\ 7 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{18}{7} \\ 0 \\ \frac{11}{7} \\ \frac{2}{7} \end{pmatrix}$$

which agrees with the original way of doing the problem.

Two good books which give more discussion of linear programming are Strang [13] and Nobel and Daniels [10]. Also listed in these books are other references which may prove useful if you are interested in seeing more on these topics. There is a great deal more which can be said about linear programming.

Spectral Theory

10.0.1 Outcomes

A. Describe the eigenvalue problem geometrically and algebraically.

B. Evaluate the spectrum and eigenvectors for a square matrix.

C. Find the principle directions of a deformation matrix.

D. Model a Markov process.

   (a) Find the limit state.

   (b) Determine comparisons of population after a long period of time.

10.1 Eigenvalues And Eigenvectors Of A Matrix

Spectral Theory refers to the study of eigenvalues and eigenvectors of a matrix. It is of fundamental importance in many areas. Row operations will no longer be such a useful tool in this subject.

10.1.1 Definition Of Eigenvectors And Eigenvalues

In this section, $\mathbb{F} = \mathbb{C}$.

To illustrate the idea behind what will be discussed, consider the following example.

Example 10.1.1 *Here is a matrix.*

$$\begin{pmatrix} 0 & 5 & -10 \\ 0 & 22 & 16 \\ 0 & -9 & -2 \end{pmatrix}.$$

Multiply this matrix by the vector

$$\begin{pmatrix} -5 \\ -4 \\ 3 \end{pmatrix}$$

and see what happens. Then multiply it by

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

and see what happens. Does this matrix act this way for some other vector?

First

$$\begin{pmatrix} 0 & 5 & -10 \\ 0 & 22 & 16 \\ 0 & -9 & -2 \end{pmatrix} \begin{pmatrix} -5 \\ -4 \\ 3 \end{pmatrix} = \begin{pmatrix} -50 \\ -40 \\ 30 \end{pmatrix} = 10 \begin{pmatrix} -5 \\ -4 \\ 3 \end{pmatrix}.$$

Next

$$\begin{pmatrix} 0 & 5 & -10 \\ 0 & 22 & 16 \\ 0 & -9 & -2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = 0 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

When you multiply the first vector by the given matrix, it stretched the vector, multiplying it by 10. When you multiplied the matrix by the second vector it sent it to the zero vector. Now consider

$$\begin{pmatrix} 0 & 5 & -10 \\ 0 & 22 & 16 \\ 0 & -9 & -2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -5 \\ 38 \\ -11 \end{pmatrix}.$$

In this case, multiplication by the matrix did not result in merely multiplying the vector by a number.

In the above example, the first two vectors were called eigenvectors and the numbers, 10 and 0 are called eigenvalues. Not every number is an eigenvalue and not every vector is an eigenvector.

Definition 10.1.2 *Let M be an $n \times n$ matrix and let $\mathbf{x} \in \mathbb{C}^n$ be a* ***nonzero vector*** *for which*

$$M\mathbf{x} = \lambda\mathbf{x} \tag{10.1}$$

for some scalar, λ. Then $\mathbf{x}$ is called an ***eigenvector*** *and λ is called an* ***eigenvalue*** *(**characteristic value**) of the matrix, M.*

Note: Eigenvectors are never equal to zero!

The set of all eigenvalues of an $n \times n$ matrix, M, is denoted by $\sigma(M)$ and is referred to as the ***spectrum*** *of M.*

The eigenvectors of a matrix M are those vectors, $\mathbf{x}$ for which multiplication by M results in a vector in the same direction or opposite direction to $\mathbf{x}$. Since the zero vector, $\mathbf{0}$ has no direction this would make no sense for the zero vector. As noted above, $\mathbf{0}$ is never allowed to be an eigenvector. How can eigenvectors be identified? Suppose $\mathbf{x}$ satisfies 10.1. Then

$$(M - \lambda I)\mathbf{x} = \mathbf{0}$$

for some $\mathbf{x} \neq \mathbf{0}$. (Equivalently, you could write $(\lambda I - M)\mathbf{x} = \mathbf{0}$.) Sometimes we will use

$$(\lambda I - M)\mathbf{x} = \mathbf{0}$$

and sometimes $(M - \lambda I)\mathbf{x} = \mathbf{0}$. It makes absolutely no difference and you should use whichever you like better. Therefore, the matrix $M - \lambda I$ cannot have an inverse because if it did, the equation could be solved,

$$\mathbf{x} = \left((M - \lambda I)^{-1}(M - \lambda I)\right)\mathbf{x} = (M - \lambda I)^{-1}((M - \lambda I)\mathbf{x}) = (M - \lambda I)^{-1}\mathbf{0} = \mathbf{0},$$

and this would require $\mathbf{x} = \mathbf{0}$, contrary to the requirement that $\mathbf{x} \neq \mathbf{0}$. By Theorem 6.2.1 on Page 85,

$$\det(M - \lambda I) = 0. \tag{10.2}$$

(Equivalently you could write $\det(\lambda I - M) = 0$.) The expression, $\det(\lambda I - M)$ or equivalently, $\det(M - \lambda I)$ is a polynomial called the **characteristic polynomial** and the above equation is called the characteristic equation. For M an $n \times n$ matrix, it follows from the theorem on expanding a matrix by its cofactor that $\det(M - \lambda I)$ is a polynomial of degree n. As such, the equation, 10.2 has a solution, $\lambda \in \mathbb{C}$ by the fundamental theorem of algebra. Is it actually an eigenvalue? The answer is yes and this follows from Observation 7.7.7 on Page 137 along with Theorem 6.2.1 on Page 85. Since $\det(M - \lambda I) = 0$ the matrix, $\det(M - \lambda I)$ cannot be one to one and so there exists a nonzero vector, $\mathbf{x}$ such that $(M - \lambda I)\mathbf{x} = \mathbf{0}$. This proves the following corollary.

Corollary 10.1.3 *Let M be an $n \times n$ matrix and $\det(M - \lambda I) = 0$. Then there exists a nonzero vector, $\mathbf{x} \in \mathbb{C}^n$ such that $(M - \lambda I)\mathbf{x} = \mathbf{0}$.*

10.1.2 Finding Eigenvectors And Eigenvalues

As an example, consider the following.

Example 10.1.4 *Find the eigenvalues and eigenvectors for the matrix,*

$$A = \begin{pmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{pmatrix}.$$

You first need to identify the eigenvalues. Recall this requires the solution of the equation

$$\det(A - \lambda I) = 0.$$

In this case this equation is

$$\det\left(\begin{pmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right) = 0$$

When you expand this determinant and simplify, you find the equation you need to solve is

$$(\lambda - 5)\left(\lambda^2 - 20\lambda + 100\right) = 0$$

and so the eigenvalues are

$$5, 10, 10.$$

We have listed 10 twice because it is a zero of multiplicity two due to

$$\lambda^2 - 20\lambda + 100 = (\lambda - 10)^2.$$

Having found the eigenvalues, it only remains to find the eigenvectors. First find the eigenvectors for $\lambda = 5$. As explained above, this requires you to solve the equation,

$$\left(\begin{pmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{pmatrix} - 5\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right)\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

That is you need to find the solution to

$$\begin{pmatrix} 0 & -10 & -5 \\ 2 & 9 & 2 \\ -4 & -8 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

By now this is an old problem. You set up the augmented matrix and row reduce to get the solution. Thus the matrix you must row reduce is

$$\left(\begin{array}{ccc|c} 0 & -10 & -5 & 0 \\ 2 & 9 & 2 & 0 \\ -4 & -8 & 1 & 0 \end{array}\right). \tag{10.3}$$

The row reduced echelon form is

$$\left(\begin{array}{ccc|c} 1 & 0 & -\frac{5}{4} & 0 \\ 0 & 1 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

and so the solution is any vector of the form

$$\begin{pmatrix} \frac{5}{4}t \\ \frac{-1}{2}t \\ t \end{pmatrix} = t \begin{pmatrix} \frac{5}{4} \\ \frac{-1}{2} \\ 1 \end{pmatrix}$$

where $t \in \mathbb{F}$. You would obtain the same collection of vectors if you replaced t with $4t$. Thus a simpler description for the solutions to this system of equations whose augmented matrix is in 10.3 is

$$t \begin{pmatrix} 5 \\ -2 \\ 4 \end{pmatrix} \tag{10.4}$$

where $t \in \mathbb{F}$. Now you need to remember that you can't take $t = 0$ because this would result in the zero vector and

Eigenvectors are never equal to zero!

Other than this value, every other choice of z in 10.4 results in an eigenvector. It is a good idea to check your work! To do so, we will take the original matrix and multiply by this vector and see if we get 5 times this vector.

$$\begin{pmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{pmatrix} \begin{pmatrix} 5 \\ -2 \\ 4 \end{pmatrix} = \begin{pmatrix} 25 \\ -10 \\ 20 \end{pmatrix} = 5 \begin{pmatrix} 5 \\ -2 \\ 4 \end{pmatrix}$$

so it appears this is correct. Always check your work on these problems if you care about getting the answer right.

The parameter, t is sometimes called a **free variable.** The set of vectors in 10.4 is called the **eigenspace** and it equals $\ker(A - \lambda I)$. You should observe that in this case the eigenspace has dimension 1 because the eigenspace is the span of a single vector. In general, you obtain the solution from the row echelon form and the number of different free variables gives you the dimension of the eigenspace. Just remember that not every vector in the eigenspace is an eigenvector. The vector, $\mathbf{0}$ is not an eigenvector although it is in the eigenspace because

Eigenvectors are never equal to zero!

Next consider the eigenvectors for $\lambda = 10$. These vectors are solutions to the equation,

$$\left(\begin{pmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{pmatrix} - 10 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

That is you must find the solutions to

$$\begin{pmatrix} -5 & -10 & -5 \\ 2 & 4 & 2 \\ -4 & -8 & -4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

which reduces to consideration of the augmented matrix,

$$\left(\begin{array}{ccc|c} -5 & -10 & -5 & 0 \\ 2 & 4 & 2 & 0 \\ -4 & -8 & -4 & 0 \end{array}\right)$$

The row reduced echelon form for this matrix is

$$\begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and so the eigenvectors are of the form

$$\begin{pmatrix} -2s - t \\ s \\ t \end{pmatrix} = s \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}.$$

You can't pick t and s both equal to zero because this would result in the zero vector and

Eigenvectors are never equal to zero!

However, every other choice of t and s does result in an eigenvector for the eigenvalue $\lambda = 10$. As in the case for $\lambda = 5$ you should check your work if you care about getting it right.

$$\begin{pmatrix} 5 & -10 & -5 \\ 2 & 14 & 2 \\ -4 & -8 & 6 \end{pmatrix} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -10 \\ 0 \\ 10 \end{pmatrix} = 10 \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$$

so it worked. The other vector will also work. Check it.

10.1.3 A Warning

The above example shows how to find eigenvectors and eigenvalues algebraically. You may have noticed it is a bit long. Sometimes students try to first row reduce the matrix before looking for eigenvalues. This is a **terrible idea** because row operations destroy the eigenvalues. The eigenvalue problem is really not about row operations.

The general eigenvalue problem is the hardest problem in algebra and people still do research on ways to find eigenvalues and their eigenvectors. If you are doing anything which would yield a way to find eigenvalues and eigenvectors for general matrices without too much trouble, the thing you are doing will certainly be wrong. The problems you will see in these notes are not too hard because they are cooked up by us to be easy. Later we will describe general methods to compute eigenvalues and eigenvectors numerically. These methods work even when the problem is not cooked up to be easy.

If you are so fortunate as to find the eigenvalues as in the above example, then finding the eigenvectors does reduce to row operations and this part of the problem is easy. However, finding the eigenvalues along with the eigenvectors is anything but easy because for an $n \times n$ matrix, it involves solving a polynomial equation of degree n. If you only find a good

approximation to the eigenvalue, it won't work. It either is or is not an eigenvalue and if it is not, the only solution to the equation, $(M-\lambda I)\,\mathbf{x}=\mathbf{0}$ will be the zero solution as explained above and

Eigenvectors are never equal to zero!

Here is another example.

Example 10.1.5 *Let*

$$A=\begin{pmatrix}2&2&-2\\1&3&-1\\-1&1&1\end{pmatrix}$$

First find the eigenvalues.

$$\det\left(\begin{pmatrix}2&2&-2\\1&3&-1\\-1&1&1\end{pmatrix}-\lambda\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}\right)=0$$

This reduces to $\lambda^3-6\lambda^2+8\lambda=0$ and the solutions are 0, 2, and 4.

0 **Can be an Eigenvalue!**

Now find the eigenvectors. For $\lambda=0$ the augmented matrix for finding the solutions is

$$\left(\begin{array}{ccc|c}2&2&-2&0\\1&3&-1&0\\-1&1&1&0\end{array}\right)$$

and the row reduced echelon form is

$$\begin{pmatrix}1&0&-1&0\\0&1&0&0\\0&0&0&0\end{pmatrix}$$

Therefore, the eigenvectors are of the form

$$t\begin{pmatrix}1\\0\\1\end{pmatrix}$$

where $t\neq 0$.

Next find the eigenvectors for $\lambda=2$. The augmented matrix for the system of equations needed to find these eigenvectors is

$$\left(\begin{array}{ccc|c}0&2&-2&0\\1&1&-1&0\\-1&1&-1&0\end{array}\right)$$

and the row reduced echelon form is

$$\begin{pmatrix}1&0&0&0\\0&1&-1&0\\0&0&0&0\end{pmatrix}$$

and so the eigenvectors are of the form

$$t\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

where $t \neq 0$.

Finally find the eigenvectors for $\lambda = 4$. The augmented matrix for the system of equations needed to find these eigenvectors is

$$\left(\begin{array}{ccc|c} -2 & 2 & -2 & 0 \\ 1 & -1 & -1 & 0 \\ -1 & 1 & -3 & 0 \end{array}\right)$$

and the row reduced echelon form is

$$\left(\begin{array}{cccc} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

Therefore, the eigenvectors are of the form

$$t\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

where $t \neq 0$.

10.1.4 Defective And Nondefective Matrices

Definition 10.1.6 *By the fundamental theorem of algebra, it is possible to write the characteristic equation in the form*

$$(\lambda - \lambda_1)^{r_1} (\lambda - \lambda_2)^{r_2} \cdots (\lambda - \lambda_m)^{r_m} = 0$$

where r_i is some integer no smaller than 1. Thus the eigenvalues are $\lambda_1, \lambda_2, \cdots, \lambda_m$. The ***algebraic multiplicity*** *of λ_j is defined to be r_j.*

Example 10.1.7 *Consider the matrix,*

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \tag{10.5}$$

What is the algebraic multiplicity of the eigenvalue $\lambda = 1$?

In this case the characteristic equation is

$$\det(A - \lambda I) = (1 - \lambda)^3 = 0$$

or equivalently,

$$\det(\lambda I - A) = (\lambda - 1)^3 = 0.$$

Therefore, λ is of algebraic multiplicity 3.

Definition 10.1.8 *The **geometric multiplicity** of an eigenvalue is the dimension of the eigenspace,*

$$\ker(A-\lambda I).$$

Example 10.1.9 *Find the geometric multiplicity of $\lambda = 1$ for the matrix in 10.5.*

We need to solve

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

The augmented matrix which must be row reduced to get this solution is therefore,

$$\left(\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

This requires $z = y = 0$ and x is arbitrary. Thus the eigenspace is

$$t\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},\ t \in \mathbb{F}.$$

It follows the geometric multiplicity of $\lambda = 1$ is 1.

Definition 10.1.10 *An $n \times n$ matrix is called **defective** if the geometric multiplicity is not equal to the algebraic multiplicity for some eigenvalue. Sometimes such an eigenvalue for which the geometric multiplicity is not equal to the algebraic multiplicity is called a defective eigenvalue. If the geometric multiplicity for an eigenvalue equals the algebraic multiplicity, the eigenvalue is sometimes referred to as nondefective.*

Here is another more interesting example of a defective matrix.

Example 10.1.11 *Let*

$$A = \begin{pmatrix} 2 & -2 & -1 \\ -2 & -1 & -2 \\ 14 & 25 & 14 \end{pmatrix}.$$

Find the eigenvectors and eigenvalues.

In this case the eigenvalues are $3, 6, 6$ where we have listed 6 twice because it is a zero of algebraic multiplicity two, the characteristic equation being

$$(\lambda - 3)(\lambda - 6)^2 = 0.$$

It remains to find the eigenvectors for these eigenvalues. First consider the eigenvectors for $\lambda = 3$. You must solve

$$\left(\begin{pmatrix} 2 & -2 & -1 \\ -2 & -1 & -2 \\ 14 & 25 & 14 \end{pmatrix} - 3\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

The augmented matrix is

$$\left(\begin{array}{ccc|c} -1 & -2 & -1 & 0 \\ -2 & -4 & -2 & 0 \\ 14 & 25 & 11 & 0 \end{array}\right)$$

and the row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

so the eigenvectors are nonzero vectors of the form

$$\begin{pmatrix} t \\ -t \\ t \end{pmatrix} = t \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$$

Next consider the eigenvectors for $\lambda = 6$. This requires you to solve

$$\left(\begin{pmatrix} 2 & -2 & -1 \\ -2 & -1 & -2 \\ 14 & 25 & 14 \end{pmatrix} - 6 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

and the augmented matrix for this system of equations is

$$\left(\begin{array}{ccc|c} -4 & -2 & -1 & 0 \\ -2 & -7 & -2 & 0 \\ 14 & 25 & 8 & 0 \end{array} \right)$$

The row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & \frac{1}{8} & 0 \\ 0 & 1 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and so the eigenvectors for $\lambda = 6$ are of the form

$$t \begin{pmatrix} -\frac{1}{8} \\ -\frac{1}{4} \\ 1 \end{pmatrix}$$

or written more simply,

$$t \begin{pmatrix} -1 \\ -2 \\ 8 \end{pmatrix}$$

where $t \in \mathbb{F}$.

Note that in this example the eigenspace for the eigenvalue, $\lambda = 6$ is of dimension 1 because there is only one parameter. However, this eigenvalue is of multiplicity two as a root to the characteristic equation. Thus this eigenvalue is a defective eigenvalue. However, the eigenvalue 3 is nondefective. The matrix is defective because it has a defective eigenvalue.

The word, defective, seems to suggest there is something wrong with the matrix. This is in fact the case. Defective matrices are a lot of trouble in applications and we may wish they never occurred. However, they do occur as the above example shows. When you study linear systems of differential equations, you will have to deal with the case of defective matrices and you will see how awful they are. The reason these matrices are so horrible to work with is that it is impossible to obtain a basis of eigenvectors. When you study differential equations, solutions to first order systems are expressed in terms of eigenvectors of a certain matrix times $e^{\lambda t}$ where λ is an eigenvalue. In order to obtain a general solution

of this sort, you must have a basis of eigenvectors. For a defective matrix, such a basis does not exist and so you have to go to something called generalized eigenvectors. Unfortunately, it is **never** explained in beginning differential equations courses why there are enough generalized eigenvectors and eigenvectors to represent the general solution. In fact, this reduces to a difficult question in linear algebra equivalent to the existence of something called the Jordan Canonical form which is much more difficult than everything discussed in the entire differential equations course. If you become interested in this, see [9]or Appendix A.

Ultimately, the algebraic issues which will occur in differential equations are a red herring anyway. The real issues relative to existence of solutions to systems of ordinary differential equations are analytical, having much more to do with calculus than with linear algebra although this will likely not be made clear when you take a beginning differential equations class.

In terms of algebra, this lack of a basis of eigenvectors says that it is impossible to obtain a diagonal matrix which is similar to the given matrix.

Although there may be repeated roots to the characteristic equation, 10.2 and it is not known whether the matrix is defective in this case, there is an important theorem which holds when considering eigenvectors which correspond to distinct eigenvalues.

Theorem 10.1.12 *Suppose* $M\mathbf{v}_i = \lambda_i \mathbf{v}_i, i = 1, \cdots, r$, $\mathbf{v}_i \neq 0$, *and that if* $i \neq j$, *then* $\lambda_i \neq \lambda_j$. *Then the set of eigenvectors,* $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ *is linearly independent.*

Proof: If the conclusion of this theorem is not true, then there exist non zero scalars, c_{k_j} such that

$$\sum_{j=1}^{m} c_{k_j} \mathbf{v}_{k_j} = \mathbf{0}. \tag{10.6}$$

Take m to be the smallest number possible for an expression of the form 10.6 to hold. Then solving for $\mathbf{v}_{k_1}$

$$\mathbf{v}_{k_1} = \sum_{k_j \neq k_1} d_{k_j} \mathbf{v}_{k_j} \tag{10.7}$$

where $d_{k_j} = c_{k_j}/c_{k_1} \neq 0$. Multiplying both sides by M,

$$\lambda_{k_1} \mathbf{v}_{k_1} = \sum_{k_j \neq k_1} d_{k_j} \lambda_{k_j} \mathbf{v}_{k_j},$$

which from 10.7 yields

$$\sum_{k_j \neq k_1} d_{k_j} \lambda_{k_1} \mathbf{v}_{k_j} = \sum_{k_j \neq k_1} d_{k_j} \lambda_{k_j} \mathbf{v}_{k_j}$$

and therefore,

$$\mathbf{0} = \sum_{k_j \neq k_1} d_{k_j} \left(\lambda_{k_1} - \lambda_{k_j}\right) \mathbf{v}_{k_j},$$

a sum having fewer than m terms. However, from the assumption that m is as small as possible for 10.6 to hold with all the scalars, c_{k_j} non zero, it follows that for some $j \neq 1$,

$$d_{k_j} \left(\lambda_{k_1} - \lambda_{k_j}\right) = 0$$

which implies $\lambda_{k_1} = \lambda_{k_j}$, a contradiction.

Here is another proof in case you did not follow the above.

Theorem 10.1.13 *Suppose* $M\mathbf{v}_i = \lambda_i \mathbf{v}_i, i = 1, \cdots, r$, $\mathbf{v}_i \neq 0$, *and that if* $i \neq j$, *then* $\lambda_i \neq \lambda_j$. *Then the set of eigenvectors,* $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ *is linearly independent.*

Proof: If the conclusion is not true, there exists a basis for span $(\mathbf{v}_1, \cdots, \mathbf{v}_r)$, consisting of some vectors from the list $\{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ which has fewer than r vectors. Let $\{\mathbf{w}_1, \cdots, \mathbf{w}_k\}$ be this list of vectors. Thus $\{\mathbf{w}_1, \cdots, \mathbf{w}_k\}$ are the pivot columns in the matrix

$$\begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_k \end{pmatrix}.$$

Then there exists $\mathbf{v} \in \{\mathbf{v}_1, \cdots, \mathbf{v}_r\}$ which is a linear combination of the $\{\mathbf{w}_1, \cdots, \mathbf{w}_k\}$. Thus

$$\mathbf{v} = \sum_{i=1}^{k} c_i \mathbf{w}_i. \tag{10.8}$$

Then doing M to both sides yields

$$\lambda_{\mathbf{v}} \mathbf{v} = \sum_{i=1}^{k} c_i \lambda_{\mathbf{w}_i} \mathbf{w}_i \tag{10.9}$$

where $\lambda_{\mathbf{v}}$ denotes the eigenvalue for $\mathbf{v}$. But also you could multiply both sides of 10.8 by $\lambda_{\mathbf{v}}$ to get

$$\lambda_{\mathbf{v}} \mathbf{v} = \sum_{i=1}^{k} c_i \lambda_{\mathbf{v}} \mathbf{w}_i.$$

And now subtracting this from 10.9 yields

$$\mathbf{0} = \sum_{i=1}^{k} c_i \left(\lambda_{\mathbf{v}} - \lambda_{\mathbf{w}_i}\right) \mathbf{w}_i$$

and by independence of the $\{\mathbf{w}_1, \cdots, \mathbf{w}_k\}$, this requires $c_i \left(\lambda_{\mathbf{v}} - \lambda_{\mathbf{w}_i}\right) = 0$ for each i. Since the eigenvalues are distinct, $\lambda_{\mathbf{v}} - \lambda_{\mathbf{w}_i} \neq 0$ and so each $c_i = 0$. But from 10.8, this requires $\mathbf{v} = \mathbf{0}$ which is impossible because $\mathbf{v}$ is an eigenvector and

Eigenvectors are never equal to zero!

This proves the theorem.

10.1.5 Complex Eigenvalues

Sometimes you have to consider eigenvalues which are complex numbers. This occurs in differential equations for example. You do these problems exactly the same way as you do the ones in which the eigenvalues are real. Here is an example.

Example 10.1.14 *Find the eigenvalues and eigenvectors of the matrix*

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 1 & 2 \end{pmatrix}.$$

You need to find the eigenvalues. Solve

$$\det\left(\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 1 & 2 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right) = 0.$$

This reduces to $(\lambda - 1)\left(\lambda^2 - 4\lambda + 5\right) = 0$. The solutions are $\lambda = 1, \lambda = 2 + i, \lambda = 2 - i$.

There is nothing new about finding the eigenvectors for $\lambda = 1$ so consider the eigenvalue $\lambda = 2 + i$. You need to solve

$$\left((2+i) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 1 & 2 \end{pmatrix} \right) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

In other words, you must consider the augmented matrix,

$$\left(\begin{array}{ccc|c} 1+i & 0 & 0 & 0 \\ 0 & i & 1 & 0 \\ 0 & -1 & i & 0 \end{array} \right)$$

for the solution. Divide the top row by $(1 + i)$ and then take $-i$ times the second row and add to the bottom. This yields

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & i & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right)$$

Now multiply the second row by $-i$ to obtain

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & -i & 0 \\ 0 & 0 & 0 & 0 \end{array} \right)$$

Therefore, the eigenvectors are of the form

$$t \begin{pmatrix} 0 \\ i \\ 1 \end{pmatrix}.$$

You should find the eigenvectors for $\lambda = 2 - i$. These are

$$t \begin{pmatrix} 0 \\ -i \\ 1 \end{pmatrix}.$$

As usual, if you want to get it right you had better check it.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ -i \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -1-2i \\ 2-i \end{pmatrix} = (2-i) \begin{pmatrix} 0 \\ -i \\ 1 \end{pmatrix}$$

so it worked.

10.2 Some Applications Of Eigenvalues And Eigenvectors

10.2.1 Principle Directions

Recall that $n \times n$ matrices can be considered as linear transformations. If F is a 3×3 real matrix having positive determinant, it can be shown that $F = RU$ where R is a rotation matrix and U is a symmetric real matrix having positive eigenvalues. An application of

this wonderful result, known to mathematicians as the **right polar factorization**, is to continuum mechanics where a chunk of material is identified with a set of points in three dimensional space.

The linear transformation, F in this context is called the **deformation gradient** and it describes the local deformation of the material. Thus it is possible to consider this deformation in terms of two processes, one which distorts the material and the other which just rotates it. It is the matrix, U which is responsible for stretching and compressing. This is why in elasticity, the stress is often taken to depend on U which is known in this context as the right **Cauchy Green strain tensor**. In this context, the eigenvalues will always be positive. The symmetry of U allows the proof of a theorem which says that if λ_M is the largest eigenvalue, then in every other direction other than the one corresponding to the eigenvector for λ_M the material is stretched less than λ_M and if λ_m is the smallest eigenvalue, then in every other direction other than the one corresponding to an eigenvector of λ_m the material is stretched more than λ_m. This process of writing a matrix as a product of two such matrices, one of which preserves distance and the other which distorts is also important in applications to geometric measure theory an interesting field of study in mathematics and to the study of quadratic forms which occur in many applications such as statistics. Here we are emphasizing the application to mechanics in which the eigenvectors of the symmetric matrix U determine the **principle directions**, those directions in which the material is stretched the most or the least.

Example 10.2.1 *Find the principle directions determined by the matrix,*

$$\begin{pmatrix} \frac{29}{11} & \frac{6}{11} & \frac{6}{11} \\ \frac{6}{11} & \frac{41}{44} & \frac{19}{44} \\ \frac{6}{11} & \frac{19}{44} & \frac{41}{44} \end{pmatrix}$$

The eigenvalues are $3, 1,$ *and* $\frac{1}{2}$.

It is nice to be given the eigenvalues. The largest eigenvalue is 3 which means that in the direction determined by the eigenvector associated with 3 the stretch is three times as large. The smallest eigenvalue is $1/2$ and so in the direction determined by the eigenvector for $1/2$ the material is stretched by a factor of $1/2$, becoming locally half as long. It remains to find these directions. First consider the eigenvector for 3. It is necessary to solve

$$\left(3\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} \frac{29}{11} & \frac{6}{11} & \frac{6}{11} \\ \frac{6}{11} & \frac{41}{44} & \frac{19}{44} \\ \frac{6}{11} & \frac{19}{44} & \frac{41}{44} \end{pmatrix} \right) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Thus the augmented matrix for this system of equations is

$$\left(\begin{array}{ccc|c} \frac{4}{11} & -\frac{6}{11} & -\frac{6}{11} & 0 \\ -\frac{6}{11} & \frac{91}{44} & -\frac{19}{44} & 0 \\ -\frac{6}{11} & -\frac{19}{44} & \frac{91}{44} & 0 \end{array}\right)$$

The row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & -3 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and so the principle direction for the eigenvalue, 3 in which the material is stretched to the maximum extent is

$$\begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}.$$

A direction vector (or unit vector) in this direction is

$$\begin{pmatrix} 3/\sqrt{11} \\ 1/\sqrt{11} \\ 1/\sqrt{11} \end{pmatrix}.$$

You should show that the direction in which the material is compressed the most is in the direction

$$\begin{pmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

Note this is meaningful information which you would have a hard time finding without the theory of eigenvectors and eigenvalues.

10.2.2 Migration Matrices

There are applications which are of great importance which feature only one eigenvalue.

Definition 10.2.2 *Let n locations be denoted by the numbers $1, 2, \cdots, n$. Also suppose it is the case that each year a_{ij} denotes the proportion of residents in location j which move to location i. Also suppose no one escapes or emigrates from without these n locations. This last assumption requires $\sum_i a_{ij} = 1$. Such matrices in which the columns are nonnegative numbers which sum to one are called* ***Markov matrices****. In this context describing migration, they are also called* ***migration matrices****.*

Example 10.2.3 *Here is an example of one of these matrices.*

$$\begin{pmatrix} .4 & .2 \\ .6 & .8 \end{pmatrix}$$

Thus if it is considered as a migration matrix, .4 is the proportion of residents in location 1 which stay in location one in a given time period while .6 is the proportion of residents in location 1 which move to location 2 and .2 is the proportion of residents in location 2 which move to location 1. Considered as a Markov matrix, these numbers are usually identified with probabilities.

If $\mathbf{v} = (x_1, \cdots, x_n)^T$ where x_i is the population of location i at a given instant, you obtain the population of location i one year later by computing $\sum_j a_{ij}x_j = (A\mathbf{v})_i$. Therefore, the population of location i after k years is $\left(A^k\mathbf{v}\right)_i$. An obvious application of this would be to a situation in which you rent trailers which can go to various parts of a city and you observe through experiments the proportion of trailers which go from point i to point j in a single day. Then you might want to find how many trailers would be in all the locations after 8 days.

Proposition 10.2.4 *Let $A = (a_{ij})$ be a migration matrix. Then 1 is always an eigenvalue for A.*

Proof: Remember that $\det\left(B^T\right) = \det(B)$. Therefore,

$$\det(A - \lambda I) = \det\left((A - \lambda I)^T\right) = \det\left(A^T - \lambda I\right)$$

because $I^T = I$. Thus the characteristic equation for A is the same as the characteristic equation for A^T and so A and A^T have the same eigenvalues. We will show that 1 is an eigenvalue for A^T and then it will follow that 1 is an eigenvalue for A.

Remember that for a migration matrix, $\sum_i a_{ij} = 1$. Therefore, if $A^T = (b_{ij})$ so $b_{ij} = a_{ji}$, it follows that

$$\sum_j b_{ij} = \sum_j a_{ji} = 1.$$

Therefore, from matrix multiplication,

$$A^T \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \sum_j b_{ij} \\ \vdots \\ \sum_j b_{ij} \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

which shows that $\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ is an eigenvector for A^T corresponding to the eigenvalue, $\lambda = 1$.

As explained above, this shows that $\lambda = 1$ is an eigenvalue for A because A and A^T have the same eigenvalues.

Example 10.2.5 *Consider the migration matrix, $\begin{pmatrix} .6 & 0 & .1 \\ .2 & .8 & 0 \\ .2 & .2 & .9 \end{pmatrix}$ for locations $1, 2,$ and 3. Suppose initially there are 100 residents in location 1, 200 in location 2 and 400 in location 4. Find the population in the three locations after 10 units of time.*

From the above, it suffices to consider

$$\begin{pmatrix} .6 & 0 & .1 \\ .2 & .8 & 0 \\ .2 & .2 & .9 \end{pmatrix}^{10} \begin{pmatrix} 100 \\ 200 \\ 400 \end{pmatrix} = \begin{pmatrix} 115.085\,829\,22 \\ 120.130\,672\,44 \\ 464.783\,498\,34 \end{pmatrix}$$

Of course you would need to round these numbers off.

A related problem asks for how many there will be in the various locations after a long time. It turns out that if some power of the migration matrix has all positive entries, then there is a limiting vector, $\mathbf{x} = \lim_{k\to\infty} A^k \mathbf{x}_0$ where $\mathbf{x}_0$ is the initial vector describing the number of inhabitants in the various locations initially. This vector will be an eigenvector for the eigenvalue 1 because

$$\mathbf{x} = \lim_{k\to\infty} A^k \mathbf{x}_0 = \lim_{k\to\infty} A^{k+1} \mathbf{x}_0 = A \lim_{k\to\infty} A^k \mathbf{x} = A\mathbf{x},$$

and the sum of its entries will equal the sum of the entries of the initial vector, $\mathbf{x}_0$ because this sum is preserved for every multiplication by A since

$$\sum_i \sum_j a_{ij} x_j = \sum_j x_j \left(\sum_i a_{ij}\right) = \sum_j x_j.$$

Here is an example. It is the same example as the one above but here it will involve the long time limit.

Example 10.2.6 *Consider the migration matrix,* $\begin{pmatrix} .6 & 0 & .1 \\ .2 & .8 & 0 \\ .2 & .2 & .9 \end{pmatrix}$ *for locations* $1, 2,$ *and* $3.$ *Suppose initially there are 100 residents in location 1, 200 in location 2 and 400 in location 4. Find the population in the three locations after a long time.*

You just need to find the eigenvector which goes with the eigenvalue 1 and then normalize it so the sum of its entries equals the sum of the entries of the initial vector. Thus you need to find a solution to

$$\left(\left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right) - \left(\begin{array}{ccc} .6 & 0 & .1 \\ .2 & .8 & 0 \\ .2 & .2 & .9 \end{array}\right)\right)\left(\begin{array}{c} x \\ y \\ z \end{array}\right) = \left(\begin{array}{c} 0 \\ 0 \\ 0 \end{array}\right)$$

The augmented matrix is

$$\left(\begin{array}{ccc|c} .4 & 0 & -.1 & 0 \\ -.2 & .2 & 0 & 0 \\ -.2 & -.2 & .1 & 0 \end{array}\right)$$

and its row reduced echelon form is

$$\left(\begin{array}{cccc} 1 & 0 & -.25 & 0 \\ 0 & 1 & -.25 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Therefore, the eigenvectors are

$$s\left(\begin{array}{c} (1/4) \\ (1/4) \\ 1 \end{array}\right)$$

and all that remains is to choose the value of s such that

$$\frac{1}{4}s + \frac{1}{4}s + s = 100 + 200 + 400$$

This yields $s = \frac{1400}{3}$ and so the long time limit would equal

$$\frac{1400}{3}\left(\begin{array}{c} (1/4) \\ (1/4) \\ 1 \end{array}\right) = \left(\begin{array}{c} 116.\,666\,666\,666\,666\,7 \\ 116.\,666\,666\,666\,666\,7 \\ 466.\,666\,666\,666\,666\,7 \end{array}\right).$$

You would of course need to round these numbers off. You see that you are not far off after just 10 units of time. Therefore, you might consider this as a useful procedure because it is probably easier to solve a simple system of equations than it is to raise a matrix to a large power.

Example 10.2.7 *Suppose a migration matrix is* $\begin{pmatrix} \frac{1}{5} & \frac{1}{2} & \frac{1}{5} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{11}{20} & \frac{1}{4} & \frac{3}{10} \end{pmatrix}$. *Find the comparison between the populations in the three locations after a long time.*

This amounts to nothing more than finding the eigenvector for $\lambda = 1$. Solve

$$\left(\left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array} \right) - \left(\begin{array}{ccc} \frac{1}{5} & \frac{1}{2} & \frac{1}{5} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{11}{20} & \frac{1}{4} & \frac{3}{10} \end{array} \right) \right) \left(\begin{array}{c} x \\ y \\ z \end{array} \right) = \left(\begin{array}{c} 0 \\ 0 \\ 0 \end{array} \right).$$

The augmented matrix is

$$\left(\begin{array}{ccc|c} \frac{4}{5} & -\frac{1}{2} & -\frac{1}{5} & 0 \\ -\frac{1}{4} & \frac{3}{4} & -\frac{1}{2} & 0 \\ -\frac{11}{20} & -\frac{1}{4} & \frac{7}{10} & 0 \end{array} \right)$$

The row echelon form is

$$\left(\begin{array}{cccc} 1 & 0 & -\frac{16}{19} & 0 \\ 0 & 1 & -\frac{18}{19} & 0 \\ 0 & 0 & 0 & 0 \end{array} \right)$$

and so an eigenvector is

$$\left(\begin{array}{c} 16 \\ 18 \\ 19 \end{array} \right).$$

Thus there will be $\frac{18}{16}^{th}$ more in location 2 than in location 1. There will be $\frac{19}{18}^{th}$ more in location 3 than in location 2.

You see the eigenvalue problem makes these sorts of determinations fairly simple.

There are many other things which can be said about these sorts of **migration problems**. They include things like the gambler's ruin problem which asks for the probability that a compulsive gambler will eventually lose all his money. However those problems are not so easy although they still involve eigenvalues and eigenvectors.

There are many other important applications of eigenvalue problems. We have just given a few such applications here. As pointed out, this is a very hard problem but sometimes you don't need to find the eigenvalues exactly.

10.3 The Estimation Of Eigenvalues

There are ways to estimate the eigenvalues for matrices from just looking at the matrix. The most famous is known as **Gerschgorin's theorem**. This theorem gives a rough idea where the eigenvalues are just from looking at the matrix.

Theorem 10.3.1 *Let A be an $n \times n$ matrix. Consider the n **Gerschgorin discs** defined as*

$$D_i \equiv \left\{ \lambda \in \mathbb{C} : |\lambda - a_{ii}| \leq \sum_{j \neq i} |a_{ij}| \right\}.$$

Then every eigenvalue is contained in some Gerschgorin disc.

This theorem says to add up the absolute values of the entries of the i^{th} row which are off the main diagonal and form the disc centered at a_{ii} having this radius. The union of these discs contains $\sigma(A)$, the spectrum of A.

Proof: Suppose $A\mathbf{x} = \lambda\mathbf{x}$ where $\mathbf{x} \neq \mathbf{0}$. Then for $A = (a_{ij})$

$$\sum_{j\neq i} a_{ij}x_j = (\lambda - a_{ii})\, x_i.$$

Therefore, picking k such that $|x_k| \geq |x_j|$ for all x_j, it follows that $|x_k| \neq 0$ since $|\mathbf{x}| \neq 0$ and

$$|x_k| \sum_{j\neq k} |a_{kj}| \geq \sum_{j\neq k} |a_{kj}|\, |x_j| \geq \left| \sum_{j\neq k} a_{kj}x_j \right| = |\lambda - a_{kk}|\, |x_k|\,.$$

Now dividing by $|x_k|$, it follows λ is contained in the k^{th} Gerschgorin disc.

Example 10.3.2 *Suppose the matrix is*

$$A = \begin{pmatrix} 21 & -16 & -6 \\ 14 & 60 & 12 \\ 7 & 8 & 38 \end{pmatrix}$$

Estimate the eigenvalues.

The exact eigenvalues are $35, 56$, and 28. The Gerschgorin disks are

$$D_1 = \{\lambda \in \mathbb{C} : |\lambda - 21| \leq 22\}\,,$$

$$D_2 = \{\lambda \in \mathbb{C} : |\lambda - 60| \leq 26\}\,,$$

and

$$D_3 = \{\lambda \in \mathbb{C} : |\lambda - 38| \leq 15\}\,.$$

Gerschgorin's theorem says these three disks contain the eigenvalues. Now 35 is in $D_3, 56$ is in D_2 and 28 is in D_1.

More can be said when the Gerschgorin disks are disjoint but this is an advanced topic which requires the theory of functions of a complex variable. If you are interested and have a background in complex variable techniques, this is in [9]

10.4 Exercises With Answers

1. Find the eigenvectors and eigenvalues of the matrix, $A = \begin{pmatrix} 8 & -3 & 1 \\ -2 & 7 & 1 \\ 0 & 0 & 10 \end{pmatrix}$. Determine whether the matrix is defective. If nondefective, diagonalize the matrix with an appropriate similarity transformation.

   First you need to write the characteristic equation.

$$\det\left(\lambda\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 8 & -3 & 1 \\ -2 & 7 & 1 \\ 0 & 0 & 10 \end{pmatrix}\right) = \det\begin{pmatrix} \lambda - 8 & 3 & -1 \\ 2 & \lambda - 7 & -1 \\ 0 & 0 & \lambda - 10 \end{pmatrix} \tag{10.10}$$

$$= \lambda^3 - 25\lambda^2 + 200\lambda - 500 = 0$$

Next you need to find the solutions to this equation. Of course this is a real joy. If there are any rational zeros they are

$$\pm\frac{\text{factor of } 500}{\text{factor of } 1}$$

I hope to find a rational zero. If there are none, then I don't know what to do at this point. This is a really lousy method for finding eigenvalues and eigenvectors. It only works if things work out well. Lets try 10. You can plug it in and see if it works or you can use synthetic division.

$$\begin{array}{cccccc} 0 & 1 & -25 & 200 & -500 \\ 10 & & 10 & -150 & 500 \\ \hline & 1 & -15 & 50 & 0 \end{array}$$

Yes, it appears 10 works and you can factor the polynomial as $(\lambda-10)\left(\lambda^2-15\lambda+50\right)$ which factors further to $(\lambda-10)(\lambda-5)(\lambda-10)$ so you find the eigenvalues are $5, 10,$ and 10. It remains to find the eigenvectors. First find an eigenvector for $\lambda=5$. To do this, you find a vector which is sent to 0 by the matrix on the right in 10.10 in which you let $\lambda=5$. Thus the augmented matrix of the system of equations you need to solve to get the eigenvector is

$$\left(\begin{array}{ccc|c} 5-8 & 3 & -1 & 0 \\ 2 & 5-7 & -1 & 0 \\ 0 & 0 & 5-10 & 0 \end{array}\right)$$

Now the row reduced echelon form is

$$\left(\begin{array}{ccc|c} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

and so you need $x=y$ and $z=0$. An eigenvector is $(1,1,0)^T$. Now you have the glorious opportunity to solve for the eigenvectors associated with $\lambda=10$. You do it the same way. The augmented matrix for the system of equations you solve to find the eigenvectors is

$$\left(\begin{array}{ccc|c} 10-8 & 3 & -1 & 0 \\ 2 & 10-7 & -1 & 0 \\ 0 & 0 & 10-10 & 0 \end{array}\right)$$

The row reduced echelon form is

$$\left(\begin{array}{ccc|c} 1 & \frac{3}{2} & -\frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

and so you need $x=-\frac{3}{2}y+\frac{1}{2}z$. It follows the eigenvectors for $\lambda=10$ are

$$\left(-\frac{3}{2}y+\frac{1}{2}z, y, z\right)^T$$

where $x, y \in \mathbb{R}$, not both equal to zero. Why? Let $y=2$ and $z=0$. This gives the vector,

$$(-3,2,0)^T$$

as one of the eigenvectors. You could also let $y = 0$ and $z = 2$ to obtain another eigenvector,

$$(1, 0, 2)^T .$$

If there exists a basis of eigenvectors, then the matrix is nondefective and as discussed above, the matrix can be diagonalized by considering $S^{-1}AS$ where the columns of S are the eigenvectors. In this case, I have found three eigenvectors and so it remains to determine whether these form a basis. Remember how to do this. You let them be the columns of a matrix and then find the rank of this matrix. If it is three, then they are a basis because they are linearly independent and the vectors are in $\mathbb{R}^3$. This is equivalent to the following matrix has an inverse.

$$\begin{pmatrix} 1 & -3 & 1 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

$$\begin{pmatrix} 1 & -3 & 1 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{2}{5} & \frac{3}{5} & -\frac{1}{5} \\ -\frac{1}{5} & \frac{1}{5} & \frac{1}{10} \\ 0 & 0 & \frac{1}{2} \end{pmatrix}$$

Then to diagonalize

$$\begin{pmatrix} \frac{2}{5} & \frac{3}{5} & -\frac{1}{5} \\ -\frac{1}{5} & \frac{1}{5} & \frac{1}{10} \\ 0 & 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 8 & -3 & 1 \\ -2 & 7 & 1 \\ 0 & 0 & 10 \end{pmatrix} \begin{pmatrix} 1 & -3 & 1 \\ 1 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} =$$

$$\begin{pmatrix} 5 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 10 \end{pmatrix}$$

Isn't this stuff marvelous! You can know this matrix is nondefective at the point when you find the eigenvectors for the repeated eigenvalue. This eigenvalue was repeated with multiplicity 2 and there were two parameters, y and z in the description of the eigenvectors. Therefore, the matrix is nondefective. Also note that there is no uniqueness for the similarity transformation.

2. Now consider the matrix, $\begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$. Find its eigenvectors and eigenvalues and determine whether it is defective.

The characteristic equation is

$$\det\left(\lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}\right) = 0$$

thus the characteristic equation is

$$(\lambda - 2)(\lambda - 1)^2 = 0.$$

The zeros are $1, 1, 2$. Lets find the eigenvectors for $\lambda = 1$. The augmented matrix for the system you need to solve is

$$\left(\begin{array}{ccc|c} -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{array}\right)$$

The row reduced echelon form is

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Then you find $x = y = 0$ and there is no restriction on z. Thus the eigenvectors are of the form

$$(0,0,z)^T,\ z \in \mathbb{R}.$$

The eigenvalue had multiplicity 2 but the eigenvectors depend on only one parameter. Therefore, the matrix is defective and cannot be diagonalized. The other eigenvector comes from row reducing the following

$$2\left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right) - \left(\begin{array}{ccc|c} 2 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \end{array}\right) = \left(\begin{array}{ccc|c} 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{array}\right)$$

The row reduced echelon form is

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Therefore the eigenvectors are of the form $(x, 0, -x)^T$. One such eigenvector is

$$(1,0,-1)^T.$$

3. Let M be an $n \times n$ matrix. Then define the adjoint of M,denoted by M^* to be the transpose of the conjugate of M. For example,

$$\left(\begin{array}{cc} 2 & i \\ 1+i & 3 \end{array}\right)^* = \left(\begin{array}{cc} 2 & 1-i \\ -i & 3 \end{array}\right).$$

A matrix, M, is self adjoint if $M^* = M$. Show the eigenvalues of a self adjoint matrix are all real. If the self adjoint matrix has all real entries, it is called symmetric. Show that the eigenvalues and eigenvectors of a symmetric matrix occur in conjugate pairs.

First note that for $\mathbf{x}$ a vector, $\mathbf{x}^*\mathbf{x} = |\mathbf{x}|^2$. This is because

$$\mathbf{x}^*\mathbf{x} = \sum_k \overline{x_k} x_k = \sum_k |x_k|^2 \equiv |\mathbf{x}|^2.$$

Also note that $(AB)^* = B^*A^*$ because this holds for transposes.This implies that for A an $n \times m$ matrix,

$$\mathbf{x}^*A^*\mathbf{x} = (A\mathbf{x})^*\mathbf{x}$$

Then if $M\mathbf{x} = \lambda\mathbf{x}$

$$\begin{aligned} \overline{\lambda}\mathbf{x}^*\mathbf{x} &= (\lambda\mathbf{x})^*\mathbf{x} = (M\mathbf{x})^*\mathbf{x} = \mathbf{x}^*M^*\mathbf{x} \\ &= \mathbf{x}^*M\mathbf{x} = \mathbf{x}^*\lambda\mathbf{x} = \lambda\mathbf{x}^*\mathbf{x} \end{aligned}$$

and so $\lambda = \overline{\lambda}$ showing that λ must be real.

4. Suppose A is an $n \times n$ matrix consisting entirely of real entries but $a+ib$ is a complex eigenvalue having the eigenvector, $\mathbf{x}+i\mathbf{y}$. Here $\mathbf{x}$ and $\mathbf{y}$ are real vectors. Show that then $a-ib$ is also an eigenvalue with the eigenvector, $\mathbf{x}-i\mathbf{y}$. **Hint:** You should remember that the conjugate of a product of complex numbers equals the product of the conjugates. Here $a+ib$ is a complex number whose conjugate equals $a-ib$.

   If A is real then the characteristic equation has all real coefficients. Therefore, letting $p(\lambda)$ be the characteristic polynomial,

$$0 = p(\lambda) = \overline{p(\lambda)} = p\left(\overline{\lambda}\right)$$

   showing that $\overline{\lambda}$ is also an eigenvalue.

5. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} -10 & -2 & 11 \\ -18 & 6 & -9 \\ 10 & -10 & -2 \end{pmatrix}.$$

   Determine whether the matrix is defective.

   The matrix has eigenvalues -12 and 18. Of these, -12 is repeated with multiplicity two. Therefore, you need to see whether the eigenspace has dimension two. If it does, then the matrix is non defective. If it does not, then the matrix is defective. The row reduced echelon form for the system you need to solve is

$$\left(\begin{array}{ccc|c} 2 & -2 & 11 & 0 \\ -18 & 18 & -9 & 0 \\ 10 & -10 & 10 & 0 \end{array}\right)$$

   and its row reduced echelon form is

$$\left(\begin{array}{ccc|c} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

   Therefore, the eigenspace is of the form

$$\begin{pmatrix} t \\ t \\ 0 \end{pmatrix}$$

   This is only one dimensional and so the matrix is defective.

6. Here is a matrix. $A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & -1 & 0 \\ 0 & -2 & 1 \end{pmatrix}$. Find a formula for A^n where n is an integer.

   First you find the eigenvectors and eigenvalues. $\begin{pmatrix} 1 & 2 & 0 \\ 0 & -1 & 0 \\ 0 & -2 & 1 \end{pmatrix}$, eigenvectors:

$$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \leftrightarrow 1, \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \leftrightarrow -1.$$

The matrix, S used to diagonalize the matrix is obtained by letting these vectors be the columns of S. Then S^{-1} is given by

$$S^{-1} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Then $S^{-1}AS$ equals

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 0 \\ 0 & -1 & 0 \\ 0 & -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \equiv D$$

Then $A = SDS^{-1}$ and $A^n = SD^nS^{-1}$. Now it is easy to find D^n.

$$D^n = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (-1)^n \end{pmatrix}$$

Therefore,

$$\begin{aligned} A^n &= \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (-1)^n \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1-(-1)^n & 0 \\ 0 & (-1)^n & 0 \\ 0 & -1+(-1)^n & 1 \end{pmatrix}. \end{aligned}$$

7. Suppose the eigenvalues of A are $\lambda_1, \cdots, \lambda_n$ and that A is nondefective. Show that $e^{At} = S \begin{pmatrix} e^{\lambda_1 t} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{\lambda_n t} \end{pmatrix} S^{-1}$ where S is the matrix which satisfies $S^{-1}AS = D$.

The diagonal matrix, D has the same characteristic equation as A why? and so it has the same eigenvalues. However the eigenvalues of D are the diagonal entries and so the diagonal entries of D are the eigenvalues of A. Now

$$S^{-1}tAS = tD$$

and

$$(tD)^n = \begin{pmatrix} (\lambda_1 t)^n & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & (\lambda_n t)^n \end{pmatrix}$$

Therefore,

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{1}{n!} (tD)^n &= \sum_{n=0}^{\infty} \frac{\left(S^{-1}tAS\right)^n}{n!} \\ &= S^{-1} \sum_{n=0}^{\infty} \frac{(tA)^n}{n!} S. \end{aligned}$$

Now the left side equals

$$\begin{aligned}\sum_{n=0}^{\infty}\frac{1}{n!}(tD)^n &= \sum_{n=0}^{\infty}\frac{1}{n!}\begin{pmatrix}(\lambda_1 t)^n & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & (\lambda_n t)^n\end{pmatrix}\\ &= \begin{pmatrix}\sum_{n=0}^{\infty}\frac{(\lambda_1 t)^n}{n!} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \sum_{n=0}^{\infty}\frac{(\lambda_n t)^n}{n!}\end{pmatrix}\\ &= \begin{pmatrix}e^{\lambda_1 t} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & e^{\lambda_n t}\end{pmatrix}.\end{aligned}$$

Therefore,

$$e^{tA}\equiv\sum_{n=0}^{\infty}\frac{(tA)^n}{n!}=S\begin{pmatrix}e^{\lambda_1 t} & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & e^{\lambda_n t}\end{pmatrix}S^{-1}.$$

Do you think you understand this? If so, think again. What exactly do you mean by an infinite sum? Actually there is no problem here. You can do this just fine and the sums converge in the sense that the ij^{th} entries converge in the partial sums. Think about this. You know what you need from calculus to see this.

8. Show that if A is similar to B then A^T is similar to B^T.

   This is easy. $A=S^{-1}BS$ and so $A^T=S^TB^T\left(S^{-1}\right)^T=S^TB^T\left(S^T\right)^{-1}$.

9. Suppose $A^m=0$ for some m a positive integer. Show that if A is diagonalizable, then $A=0$.

   Since $A^m=0$ suppose $S^{-1}AS=D$. Then raising to the m^{th} power, $D^m=S^{-1}A^mS=0$. Therefore, $D=0$. But then $A=S0S^{-1}=0$.

10. Find the complex eigenvalues and eigenvectors of the matrix $\begin{pmatrix}1 & 1 & -6\\ 7 & -5 & -6\\ -1 & 7 & 2\end{pmatrix}$.

    Determine whether the matrix is defective.

    After wading through much affliction you find the eigenvalues are $-6, 2+6i, 2-6i$. Since these are distinct, the matrix cannot be defective. We must find the eigenvectors for these eigenvalues. The augmented matrix for the system of equations which must be solved to find the eigenvectors associated with $2-6i$ is

$$\left(\begin{array}{ccc|c}-1+6i & 1 & -6 & 0\\ 7 & -7+6i & -6 & 0\\ -1 & 7 & 6i & 0\end{array}\right).$$

    The row reduced echelon form is

$$\begin{pmatrix}1 & 0 & i & 0\\ 0 & 1 & i & 0\\ 0 & 0 & 0 & 0\end{pmatrix}$$

and so the eigenvectors are of the form

$$t\begin{pmatrix} -i \\ -i \\ 1 \end{pmatrix}.$$

You can check this as follows

$$\begin{pmatrix} 1 & 1 & -6 \\ 7 & -5 & -6 \\ -1 & 7 & 2 \end{pmatrix}\begin{pmatrix} -i \\ -i \\ 1 \end{pmatrix} = \begin{pmatrix} -6-2i \\ -6-2i \\ 2-6i \end{pmatrix}$$

and

$$(2-6i)\begin{pmatrix} -i \\ -i \\ 1 \end{pmatrix} = \begin{pmatrix} -6-2i \\ -6-2i \\ 2-6i \end{pmatrix}.$$

It follows that the eigenvectors for $\lambda = 2+6i$ are

$$t\begin{pmatrix} i \\ i \\ 1 \end{pmatrix}.$$

This is because A is real. If $A\mathbf{v} = \lambda\mathbf{v}$, then taking the conjugate,

$$A\overline{\mathbf{v}} = \overline{A\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}.$$

It only remains to find the eigenvector for $\lambda = -6$. The augmented matrix to row reduce is

$$\left(\begin{array}{ccc|c} 7 & 1 & -6 & 0 \\ 7 & 1 & -6 & 0 \\ -1 & 7 & 8 & 0 \end{array}\right)$$

The row reduced echelon form is

$$\left(\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

Then an eigenvector is

$$\begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}.$$

Some Special Matrices

11.0.1 Outcomes

A. Define symmetric matrix, skew-symmetric matrix, and orthogonal matrix. Prove identities involving these types of matrices.

B. Characterize and determine the eigenvalues and eigenvectors of symmetric, skew-symmetric, and orthogonal matrices. Derive basic facts concerning these matrices.

C. Define an orthonormal set of vectors. Determine whether a set of vectors is orthonormal.

D. Relate the orthogonality of a matrix to the orthonormality of its column (or row) vectors.

E. Diagonalize a symmetric matrix. In particular, given a symmetric matrix, A, find an orthogonal matrix, U and a diagonal matrix, D such that $U^T AU = D$.

F. Understand and use the Gram Schmidt process.

G. Understand and use the technique of least square approximations.

11.1 Symmetric And Orthogonal Matrices

11.1.1 Orthogonal Matrices

Remember that to find the inverse of a matrix was often a long process. However, it was very easy to take the transpose of a matrix. For some matrices, the transpose equals the inverse and when the matrix has all real entries, and this is true, it is called an orthogonal matrix.

Definition 11.1.1 *A real $n \times n$ matrix, U is called an **Orthogonal** matrix if $UU^T = U^T U = I$.*

Example 11.1.2 *Show the matrix,*

$$U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}$$

is orthogonal.

$$UU^T = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Example 11.1.3 *Let* $U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}$. *Is* U *orthogonal?*

The answer is yes. This is because the columns form an orthonormal set of vectors as well as the rows. As discussed above this is equivalent to $U^TU = I$.

$$U^TU = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}^T \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

When you say that U is orthogonal, you are saying that

$$\sum_j U_{ij}U^T_{jk} = \sum_j U_{ij}U_{kj} = \delta_{ik}.$$

In words, the dot product of the i^{th} row of U with the k^{th} row gives 1 if $i = k$ and 0 if $i \neq k$. The same is true of the columns because $U^TU = I$ also. Therefore,

$$\sum_j U^T_{ij}U_{jk} = \sum_j U_{ji}U_{jk} = \delta_{ik}$$

which says that the one column dotted with another column gives 1 if the two columns are the same and 0 if the two columns are different.

More succinctly, this states that if $\mathbf{u}_1, \cdots, \mathbf{u}_n$ are the columns of U, an orthogonal matrix, then

$$\mathbf{u}_i \cdot \mathbf{u}_j = \delta_{ij} \equiv \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}. \tag{11.1}$$

Definition 11.1.4 *A set of vectors,* $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ *is said to be an* ***orthonormal*** *set if 11.1.*

Theorem 11.1.5 *If* $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ *is an orthonormal set of vectors then it is linearly independent.*

Proof: Using the properties of the dot product,

$$\mathbf{0} \cdot \mathbf{u} = (\mathbf{0} + \mathbf{0}) \cdot \mathbf{u} = \mathbf{0} \cdot \mathbf{u} + \mathbf{0} \cdot \mathbf{u}$$

and so, subtracting $\mathbf{0} \cdot \mathbf{u}$ from both sides yields $\mathbf{0} \cdot \mathbf{u} = 0$. Now suppose $\sum_j c_j\mathbf{u}_j = \mathbf{0}$. Then from the properties of the dot product,

$$c_k = \sum_j c_j\delta_{jk} = \sum_j c_j\left(\mathbf{u}_j \cdot \mathbf{u}_k\right) = \left(\sum_j c_j\mathbf{u}_j\right) \cdot \mathbf{u}_k = \mathbf{0} \cdot \mathbf{u}_k = 0.$$

Since k was arbitrary, this shows that each $c_k = 0$ and this has shown that if $\sum_j c_j\mathbf{u}_j = \mathbf{0}$, then each $c_j = 0$. This is what it means for the set of vectors to be linearly independent.

Example 11.1.6 *Let* $U = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & 0 & -\frac{\sqrt{6}}{3} \end{pmatrix}$. *Is* U *an orthogonal matrix?*

The answer is yes. This is because the columns (rows) form an orthonormal set of vectors.

The importance of orthogonal matrices is that they change components of vectors relative to different Cartesian coordinate systems. Geometrically, the orthogonal matrices are exactly those which preserve all distances in the sense that if $\mathbf{x} \in \mathbb{R}^n$ and U is orthogonal, then $||U\mathbf{x}|| = ||\mathbf{x}||$ because

$$||U\mathbf{x}||^2 = (U\mathbf{x})^T U\mathbf{x} = \mathbf{x}^T U^T U\mathbf{x} = \mathbf{x}^T I\mathbf{x} = ||\mathbf{x}||^2 .$$

Observation 11.1.7 *Suppose* U *is an orthogonal matrix. Then* $\det(U) = \pm 1$.

This is easy to see from the properties of determinants. Thus

$$\det(U)^2 = \det\left(U^T\right)\det(U) = \det\left(U^T U\right) = \det(I) = 1.$$

Orthogonal matrices are divided into two classes, proper and improper. The proper orthogonal matrices are those whose determinant equals 1 and the improper ones are those whose determinants equal -1. The reason for the distinction is that the improper orthogonal matrices are sometimes considered to have no physical significance since they cause a change in orientation which would correspond to material passing through itself in a non physical manner. Thus in considering which coordinate systems must be considered in certain applications, you only need to consider those which are related by a proper orthogonal transformation. Geometrically, the linear transformations determined by the proper orthogonal matrices correspond to the composition of rotations.

11.1.2 Symmetric And Skew Symmetric Matrices

Definition 11.1.8 *A real* $n \times n$ *matrix,* A, *is* ***symmetric*** *if* $A^T = A$. *If* $A = -A^T$, *then* A *is called* ***skew symmetric.***

Theorem 11.1.9 *The eigenvalues of a real symmetric matrix are real. The eigenvalues of a real skew symmetric matrix are 0 or pure imaginary.*

Proof: The proof of this theorem is in [9]. It is best understood as a special case of more general considerations. However, here is a proof in this special case.

Recall that for a complex number, $a + ib$, the complex conjugate, denoted by $\overline{a+ib}$ is given by the formula $\overline{a+ib} = a - ib$. The notation, $\overline{\mathbf{x}}$ will denote the vector which has every entry replaced by its complex conjugate.

Suppose A is a real symmetric matrix and $A\mathbf{x} = \lambda\mathbf{x}$. Then

$$\overline{\lambda}\overline{\mathbf{x}}^T\mathbf{x} = \left(\overline{A\mathbf{x}}\right)^T\mathbf{x} = \overline{\mathbf{x}}^T A^T\mathbf{x} = \overline{\mathbf{x}}^T A\mathbf{x} = \lambda\overline{\mathbf{x}}^T\mathbf{x}.$$

Dividing by $\overline{\mathbf{x}}^T\mathbf{x}$ on both sides yields $\overline{\lambda} = \lambda$ which says λ is real. (Why?)

Next suppose $A = -A^T$ so A is skew symmetric and $A\mathbf{x} = \lambda\mathbf{x}$. Then

$$\overline{\lambda}\overline{\mathbf{x}}^T\mathbf{x} = \left(\overline{A\mathbf{x}}\right)^T\mathbf{x} = \overline{\mathbf{x}}^T A^T\mathbf{x} = -\overline{\mathbf{x}}^T A\mathbf{x} = -\lambda\overline{\mathbf{x}}^T\mathbf{x}$$

and so, dividing by $\overline{\mathbf{x}}^T\mathbf{x}$ as before, $\overline{\lambda} = -\lambda$. Letting $\lambda = a + ib$, this means $a - ib = -a - ib$ and so $a = 0$. Thus λ is pure imaginary.

Example 11.1.10 *Let* $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. *This is a skew symmetric matrix. Find its eigenvalues.*

Its eigenvalues are obtained by solving the equation $\det \begin{pmatrix} -\lambda & -1 \\ 1 & -\lambda \end{pmatrix} = \lambda^2 + 1 = 0$. You see the eigenvalues are $\pm i$, pure imaginary.

Example 11.1.11 *Let* $A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$. *This is a symmetric matrix. Find its eigenvalues.*

Its eigenvalues are obtained by solving the equation, $\det \begin{pmatrix} 1-\lambda & 2 \\ 2 & 3-\lambda \end{pmatrix} = -1 - 4\lambda + \lambda^2 = 0$ and the solution is $\lambda = 2 + \sqrt{5}$ and $\lambda = 2 - \sqrt{5}$.

Definition 11.1.12 *An* $n \times n$ *matrix,* $A = (a_{ij})$ *is called a* ***diagonal matrix*** *if* $a_{ij} = 0$ *whenever* $i \neq j$. *For example, a diagonal matrix is of the form indicated below where* $*$ *denotes a number.*

$$\begin{pmatrix} * & 0 & \cdots & 0 \\ 0 & * & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & * \end{pmatrix}$$

Theorem 11.1.13 *Let* A *be a real symmetric matrix. Then there exists an orthogonal matrix,* U *such that* $U^T AU$ *is a diagonal matrix. Moreover, the diagonal entries are the eigenvalues of* A.

Proof: The proof may be found in [9].

Corollary 11.1.14 *If* A *is a real* $n \times n$ *symmetric matrix, then there exists an orthonormal set of eigenvectors,* $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$.

Proof: Since A is symmetric, then by Theorem 11.1.13, there exists an orthogonal matrix, U such that $U^T AU = D$, a diagonal matrix whose diagonal entries are the eigenvalues of A. Therefore, since A is symmetric and all the matrices are real,

$$\overline{D} = \overline{D^T} = \overline{U^T A^T U} = U^T A^T U = U^T AU = D$$

showing D is real because each entry of D equals its complex conjugate.[1]

Finally, let

$$U = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$$

where the $\mathbf{u}_i$ denote the columns of U and

$$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

The equation, $U^T AU = D$ implies

$$\begin{aligned} AU &= \begin{pmatrix} A\mathbf{u}_1 & A\mathbf{u}_2 & \cdots & A\mathbf{u}_n \end{pmatrix} \\ &= UD = \begin{pmatrix} \lambda_1 \mathbf{u}_1 & \lambda_2 \mathbf{u}_2 & \cdots & \lambda_n \mathbf{u}_n \end{pmatrix} \end{aligned}$$

[1] Recall that for a complex number, $x + iy$, the complex conjugate, denoted by $\overline{x+iy}$ is defined as $x - iy$.

where the entries denote the columns of AU and UD respectively. Therefore, $A\mathbf{u}_i = \lambda_i \mathbf{u}_i$ and since the matrix is orthogonal, the ij^{th} entry of $U^T U$ equals δ_{ij} and so

$$\delta_{ij} = \mathbf{u}_i^T \mathbf{u}_j = \mathbf{u}_i \cdot \mathbf{u}_j.$$

This proves the corollary because it shows the vectors $\{\mathbf{u}_i\}$ form an orthonormal basis.

The following corollary is also important.

Example 11.1.15 *Find the eigenvalues and an orthonormal basis of eigenvectors for the matrix,*

$$\begin{pmatrix} \frac{19}{9} & -\frac{8}{15}\sqrt{5} & \frac{2}{45}\sqrt{5} \\ -\frac{8}{15}\sqrt{5} & -\frac{1}{5} & -\frac{16}{15} \\ \frac{2}{45}\sqrt{5} & -\frac{16}{15} & \frac{94}{45} \end{pmatrix}$$

given that the eigenvalues are 3, -1, and 2.

The augmented matrix which needs to be row reduced to find the eigenvectors for $\lambda = 3$ is

$$\left(\begin{array}{ccc|c} \frac{19}{9} - 3 & -\frac{8}{15}\sqrt{5} & \frac{2}{45}\sqrt{5} & 0 \\ -\frac{8}{15}\sqrt{5} & -\frac{1}{5} - 3 & -\frac{16}{15} & 0 \\ \frac{2}{45}\sqrt{5} & -\frac{16}{15} & \frac{94}{45} - 3 & 0 \end{array}\right)$$

and the row reduced echelon form for this is

$$\left(\begin{array}{ccc|c} 1 & 0 & -\frac{1}{2}\sqrt{5} & 0 \\ 0 & 1 & \frac{3}{4} & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Therefore, eigenvectors for $\lambda = 3$ are

$$z \begin{pmatrix} \frac{1}{2}\sqrt{5} \\ -\frac{3}{4} \\ 1 \end{pmatrix}$$

where $z \neq 0$.

The augmented matrix which must be row reduced to find the eigenvectors for $\lambda = -1$ is

$$\left(\begin{array}{ccc|c} \frac{19}{9} + 1 & -\frac{8}{15}\sqrt{5} & \frac{2}{45}\sqrt{5} & 0 \\ -\frac{8}{15}\sqrt{5} & -\frac{1}{5} + 1 & -\frac{16}{15} & 0 \\ \frac{2}{45}\sqrt{5} & -\frac{16}{15} & \frac{94}{45} + 1 & 0 \end{array}\right)$$

and the row reduced echelon form is

$$\left(\begin{array}{ccc|c} 1 & 0 & -\frac{1}{2}\sqrt{5} & 0 \\ 0 & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

Therefore, the eigenvectors for $\lambda = -1$ are

$$z \begin{pmatrix} \frac{1}{2}\sqrt{5} \\ 3 \\ 1 \end{pmatrix}, \ z \neq 0$$

The augmented matrix which must be row reduced to find the eigenvectors for $\lambda = 2$ is

$$\begin{pmatrix} \frac{19}{9} - 2 & -\frac{8}{15}\sqrt{5} & \frac{2}{45}\sqrt{5} & | & 0 \\ -\frac{8}{15}\sqrt{5} & -\frac{1}{5} - 2 & -\frac{16}{15} & | & 0 \\ \frac{2}{45}\sqrt{5} & -\frac{16}{15} & \frac{94}{45} - 2 & | & 0 \end{pmatrix}$$

and its row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & \frac{2}{5}\sqrt{5} & | & 0 \\ 0 & 1 & 0 & | & 0 \\ 0 & 0 & 0 & | & 0 \end{pmatrix}$$

so the eigenvectors for $\lambda = 2$ are

$$z \begin{pmatrix} -\frac{2}{5}\sqrt{5} \\ 0 \\ 1 \end{pmatrix}, \ z \neq 0.$$

It remains to find an orthonormal basis. You can check that the dot product of any of these vectors with another of them gives zero and so it suffices choose z in each case such that the resulting vector has length 1. First consider the vectors for $\lambda = 3$. It is required to choose z such that

$$z \begin{pmatrix} \frac{1}{2}\sqrt{5} \\ -\frac{3}{4} \\ 1 \end{pmatrix}$$

is a unit vector. In other words, you need

$$z \begin{pmatrix} \frac{1}{2}\sqrt{5} \\ -\frac{3}{4} \\ 1 \end{pmatrix} \cdot z \begin{pmatrix} \frac{1}{2}\sqrt{5} \\ -\frac{3}{4} \\ 1 \end{pmatrix} = 1.$$

But the above dot product equals $\frac{45}{16}z^2$ and this equals 1 when $z = \frac{4}{15}\sqrt{5}$. Therefore, the eigenvector which is desired is

$$\frac{4}{15}\sqrt{5} \begin{pmatrix} \frac{1}{2}\sqrt{5} \\ -\frac{3}{4} \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} \\ -\frac{1}{5}\sqrt{5} \\ \frac{4}{15}\sqrt{5} \end{pmatrix}.$$

Next find the eigenvector for $\lambda = -1$. The same process requires that $1 = \frac{45}{4}z^2$ which happens when $z = \frac{2}{15}\sqrt{5}$. Therefore, an eigenvector for $\lambda = -1$ which has unit length is

$$\frac{2}{15}\sqrt{5}\begin{pmatrix} \frac{1}{2}\sqrt{5} \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{2}{5}\sqrt{5} \\ \frac{2}{15}\sqrt{5} \end{pmatrix}.$$

Finally, consider $\lambda = 2$. This time you need $1 = \frac{9}{5}z^2$ which occurs when $z = \frac{1}{3}\sqrt{5}$. Therefore, the eigenvector is

$$\frac{1}{3}\sqrt{5}\begin{pmatrix} -\frac{2}{5}\sqrt{5} \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -\frac{2}{3} \\ 0 \\ \frac{1}{3}\sqrt{5} \end{pmatrix}.$$

Now recall that the vectors form an orthonormal set of vectors if the matrix having them as columns is orthogonal. That matrix is

$$\begin{pmatrix} \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ -\frac{1}{5}\sqrt{5} & \frac{2}{5}\sqrt{5} & 0 \\ \frac{4}{15}\sqrt{5} & \frac{2}{15}\sqrt{5} & \frac{1}{3}\sqrt{5} \end{pmatrix}.$$

Is this orthogonal? To find out, multiply by its transpose. Thus

$$\begin{pmatrix} \frac{2}{3} & -\frac{1}{5}\sqrt{5} & \frac{4}{15}\sqrt{5} \\ \frac{1}{3} & \frac{2}{5}\sqrt{5} & \frac{2}{15}\sqrt{5} \\ -\frac{2}{3} & 0 & \frac{1}{3}\sqrt{5} \end{pmatrix}\begin{pmatrix} \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ -\frac{1}{5}\sqrt{5} & \frac{2}{5}\sqrt{5} & 0 \\ \frac{4}{15}\sqrt{5} & \frac{2}{15}\sqrt{5} & \frac{1}{3}\sqrt{5} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Since the identity was obtained this shows the above matrix is orthogonal and that therefore, the columns form an orthonormal set of vectors. The problem asks for you to find an orthonormal basis. However, you can show that an orthonormal set of n vectors in $\mathbb{R}^n$ is always a basis. Therefore, since there are three of these vectors, they must constitute a basis.

Example 11.1.16 *Find an orthonormal set of three eigenvectors for the matrix,*

$$\begin{pmatrix} \frac{13}{9} & \frac{2}{15}\sqrt{5} & \frac{8}{45}\sqrt{5} \\ \frac{2}{15}\sqrt{5} & \frac{6}{5} & \frac{4}{15} \\ \frac{8}{45}\sqrt{5} & \frac{4}{15} & \frac{61}{45} \end{pmatrix}$$

given the eigenvalues are 2, *and* 1.

The eigenvectors which go with $\lambda = 2$ are obtained from row reducing the matrix

$$\left(\begin{array}{ccc|c} \frac{13}{9} - 2 & \frac{2}{15}\sqrt{5} & \frac{8}{45}\sqrt{5} & 0 \\ \frac{2}{15}\sqrt{5} & \frac{6}{5} - 2 & \frac{4}{15} & 0 \\ \frac{8}{45}\sqrt{5} & \frac{4}{15} & \frac{61}{45} - 2 & 0 \end{array}\right)$$

and its row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & -\frac{1}{2}\sqrt{5} & | & 0 \\ 0 & 1 & -\frac{3}{4} & | & 0 \\ 0 & 0 & 0 & | & 0 \end{pmatrix}$$

which shows the eigenvectors for $\lambda = 2$ are

$$z \begin{pmatrix} \frac{1}{2}\sqrt{5} \\ \frac{3}{4} \\ 1 \end{pmatrix}$$

and a choice for z which will produce a unit vector is $z = \frac{4}{15}\sqrt{5}$. Therefore, the vector we want is

$$\begin{pmatrix} \frac{2}{3} \\ \frac{1}{5}\sqrt{5} \\ \frac{4}{15}\sqrt{5} \end{pmatrix}.$$

Next consider the eigenvectors for $\lambda = 1$. The matrix which must be row reduced is

$$\begin{pmatrix} \frac{13}{9} - 1 & \frac{2}{15}\sqrt{5} & \frac{8}{45}\sqrt{5} & | & 0 \\ \frac{2}{15}\sqrt{5} & \frac{6}{5} - 1 & \frac{4}{15} & | & 0 \\ \frac{8}{45}\sqrt{5} & \frac{4}{15} & \frac{61}{45} - 1 & | & 0 \end{pmatrix}$$

and its row reduced echelon form is

$$\begin{pmatrix} 1 & \frac{3}{10}\sqrt{5} & \frac{2}{5}\sqrt{5} & | & 0 \\ 0 & 0 & 0 & | & 0 \\ 0 & 0 & 0 & | & 0 \end{pmatrix}.$$

Therefore, the eigenvectors are of the form

$$\begin{pmatrix} -\frac{3}{10}\sqrt{5}y - \frac{2}{5}\sqrt{5}z \\ y \\ z \end{pmatrix}.$$

This is a two dimensional eigenspace.

Before going further, we want to point out that no matter how we choose y and z the resulting vector will be orthogonal to the eigenvector for $\lambda = 2$. This is a special case of a general result which states that eigenvectors for distinct eigenvalues of a symmetric matrix are orthogonal. For this case you need to show the following dot product equals zero.

$$\begin{pmatrix} \frac{2}{3} \\ \frac{1}{5}\sqrt{5} \\ \frac{4}{15}\sqrt{5} \end{pmatrix} \cdot \begin{pmatrix} -\frac{3}{10}\sqrt{5}y - \frac{2}{5}\sqrt{5}z \\ y \\ z \end{pmatrix} \tag{11.2}$$

This is left for you to do.

Continuing with the task of finding an orthonormal basis, Let $y = 0$ first. This results in eigenvectors of the form

$$\begin{pmatrix} -\frac{2}{5}\sqrt{5}z \\ 0 \\ z \end{pmatrix}$$

and letting $z = \frac{1}{3}\sqrt{5}$ you obtain a unit vector. Thus the second vector will be

$$\begin{pmatrix} -\frac{2}{5}\sqrt{5}\left(\frac{1}{3}\sqrt{5}\right) \\ 0 \\ \frac{1}{3}\sqrt{5} \end{pmatrix} = \begin{pmatrix} -\frac{2}{3} \\ 0 \\ \frac{1}{3}\sqrt{5} \end{pmatrix}.$$

It remains to find the third vector in the orthonormal basis. This merely involves choosing y and z in 11.2 in such a way that the resulting vector has dot product with the two given vectors equal to zero. Thus you need

$$\begin{pmatrix} -\frac{3}{10}\sqrt{5}y - \frac{2}{5}\sqrt{5}z \\ y \\ z \end{pmatrix} \cdot \begin{pmatrix} -\frac{2}{3} \\ 0 \\ \frac{1}{3}\sqrt{5} \end{pmatrix} = \frac{1}{5}\sqrt{5}y + \frac{3}{5}\sqrt{5}z = 0.$$

The dot product with the eigenvector for $\lambda = 2$ is automatically equal to zero and so all that you need is the above equation. This is satisfied when $z = -\frac{1}{3}y$. Therefore, the vector we want is of the form

$$\begin{pmatrix} -\frac{3}{10}\sqrt{5}y - \frac{2}{5}\sqrt{5}\left(-\frac{1}{3}y\right) \\ y \\ \left(-\frac{1}{3}y\right) \end{pmatrix} = \begin{pmatrix} -\frac{1}{6}\sqrt{5}y \\ y \\ -\frac{1}{3}y \end{pmatrix}$$

and it only remains to choose y in such a way that this vector has unit length. This occurs when $y = \frac{2}{5}\sqrt{5}$. Therefore, the vector we want is

$$\frac{2}{5}\sqrt{5}\begin{pmatrix} -\frac{1}{6}\sqrt{5} \\ 1 \\ -\frac{1}{3} \end{pmatrix} = \begin{pmatrix} -\frac{1}{3} \\ \frac{2}{5}\sqrt{5} \\ -\frac{2}{15}\sqrt{5} \end{pmatrix}.$$

The three eigenvectors which constitute an orthonormal basis are

$$\begin{pmatrix} -\frac{1}{3} \\ \frac{2}{5}\sqrt{5} \\ -\frac{2}{15}\sqrt{5} \end{pmatrix}, \begin{pmatrix} -\frac{2}{3} \\ 0 \\ \frac{1}{3}\sqrt{5} \end{pmatrix}, \text{ and } \begin{pmatrix} \frac{2}{3} \\ \frac{1}{5}\sqrt{5} \\ \frac{4}{15}\sqrt{5} \end{pmatrix}.$$

To check our work and see if this is really an orthonormal set of vectors, we make them the columns of a matrix and see if the resulting matrix is orthogonal. The matrix is

$$\begin{pmatrix} -\frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{5}\sqrt{5} & 0 & \frac{1}{5}\sqrt{5} \\ -\frac{2}{15}\sqrt{5} & \frac{1}{3}\sqrt{5} & \frac{4}{15}\sqrt{5} \end{pmatrix}.$$

This matrix times its transpose equals

$$\begin{pmatrix} -\frac{1}{3} & -\frac{2}{3} & \frac{2}{3} \\ \frac{2}{5}\sqrt{5} & 0 & \frac{1}{5}\sqrt{5} \\ -\frac{2}{15}\sqrt{5} & \frac{1}{3}\sqrt{5} & \frac{4}{15}\sqrt{5} \end{pmatrix} \begin{pmatrix} -\frac{1}{3} & \frac{2}{5}\sqrt{5} & -\frac{2}{15}\sqrt{5} \\ -\frac{2}{3} & 0 & \frac{1}{3}\sqrt{5} \\ \frac{2}{3} & \frac{1}{5}\sqrt{5} & \frac{4}{15}\sqrt{5} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and so this is indeed an orthonormal basis.

Because of the repeated eigenvalue, there would have been many other orthonormal bases which could have been obtained. It was pretty arbitrary for to take $y = 0$ in the above argument. We could just as easily have taken $z = 0$ or even $y = z = 1$. Any such change would have resulted in a different orthonormal basis. Geometrically, what is happening is the eigenspace for $\lambda = 1$ was two dimensional. It can be visualized as a plane in three dimensional space which passes through the origin. There are infinitely many different pairs of perpendicular unit vectors in this plane.

11.1.3 Diagonalizing A Symmetric Matrix

Recall the following definition:

Definition 11.1.17 *An $n \times n$ matrix, $A = (a_{ij})$ is called a diagonal matrix if $a_{ij} = 0$ whenever $i \neq j$. For example, a diagonal matrix is of the form indicated below where $*$ denotes a number.*

$$\begin{pmatrix} * & 0 & \cdots & 0 \\ 0 & * & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & * \end{pmatrix}$$

Definition 11.1.18 *An $n \times n$ matrix, A is said to be **non defective** or **diagonalizable** if there exists an invertible matrix, S such that $S^{-1}AS = D$ where D is a diagonal matrix as described above.*

Some matrices are non defective and some are not. As indicated in Theorem 11.1.13 if A is a real symmetric matrix, there exists an orthogonal matrix, U such that $U^TAU = D$ a diagonal matrix. Therefore, every symmetric matrix is non defective because if U is an orthogonal matrix, its inverse is U^T. In the following example, this orthogonal matrix will be found.

Example 11.1.19 *Let $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{3}{2} \end{pmatrix}$. Find an orthogonal matrix, U such that U^TAU is a diagonal matrix.*

In this case, a tedious computation shows the eigenvalues are 2 and 1. First we will find an eigenvector for the eigenvalue 2. This involves row reducing the following augmented matrix.

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 2-\frac{3}{2} & -\frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 2-\frac{3}{2} & 0 \end{array}\right)$$

The row reduced echelon form is

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

and so an eigenvector is

$$\left(\begin{array}{c} 0 \\ 1 \\ 1 \end{array}\right).$$

However, it is desired that the eigenvectors obtained all be unit vectors and so dividing this vector by its length gives

$$\left(\begin{array}{c} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{array}\right).$$

Next consider the case of the eigenvalue, 1. The matrix which needs to be row reduced in this case is

$$\left(\begin{array}{ccc|c} 0 & 0 & 0 & 0 \\ 0 & 1-\frac{3}{2} & -\frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 1-\frac{3}{2} & 0 \end{array}\right)$$

The row reduced echelon form is

$$\left(\begin{array}{ccc|c} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

Therefore, the eigenvectors are of the form

$$\left(\begin{array}{c} s \\ -t \\ t \end{array}\right).$$

Two of these which are orthonormal are

$$\left(\begin{array}{c} 1 \\ 0 \\ 0 \end{array}\right) \text{ and } \left(\begin{array}{c} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{array}\right).$$

An orthogonal matrix which works in the process is then obtained by letting these vectors be the columns.

$$\left(\begin{array}{ccc} 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{array}\right).$$

It remains to verify this works. U^TAU is of the form

$$\left(\begin{array}{ccc} 0 & -\frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \\ 1 & 0 & 0 \\ 0 & \frac{1}{2}\sqrt{2} & \frac{1}{2}\sqrt{2} \end{array}\right)\left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & \frac{3}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{3}{2} \end{array}\right)\left(\begin{array}{ccc} 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{array}\right)$$

$$= \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{array}\right),$$

the desired diagonal matrix.

11.2 Fundamental Theory And Generalizations*

11.2.1 Block Multiplication Of Matrices

Consider the following problem

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} E & F \\ G & H \end{pmatrix}$$

You know how to do this. You get

$$\begin{pmatrix} AE+BG & AF+BH \\ CE+DG & CF+DH \end{pmatrix}.$$

Now what if instead of numbers, the entries, A, B, C, D, E, F, G are matrices of a size such that the multiplications and additions needed in the above formula all make sense. Would the formula be true in this case? I will show below that this is true.

Suppose A is a matrix of the form

$$A = \begin{pmatrix} A_{11} & \cdots & A_{1m} \\ \vdots & \ddots & \vdots \\ A_{r1} & \cdots & A_{rm} \end{pmatrix} \tag{11.3}$$

where A_{ij} is a $s_i \times p_j$ matrix where s_i is constant for $j = 1, \cdots, m$ for each $i = 1, \cdots, r$. Such a matrix is called a **block matrix,** also a **partitioned matrix**. How do you get the block A_{ij}? Here is how for A an $m \times n$ matrix:

$$\overbrace{\begin{pmatrix} 0 & I_{s_i \times s_i} & 0 \end{pmatrix}}^{s_i \times m} A \overbrace{\begin{pmatrix} 0 \\ I_{p_j \times p_j} \\ 0 \end{pmatrix}}^{n \times p_j}. \tag{11.4}$$

In the block column matrix on the right, you need to have $c_j - 1$ rows of zeros above the small $p_j \times p_j$ identity matrix where the columns of A involved in A_{ij} are $c_j, \cdots, c_j + p_j$ and in the block row matrix on the left, you need to have $r_i - 1$ columns of zeros to the left of the $s_i \times s_i$ identity matrix where the rows of A involved in A_{ij} are $r_i, \cdots, r_i + s_i$. An important observation to make is that the matrix on the right specifies columns to use in the block and the one on the left specifies the rows used. There is no overlap between the blocks of A. Thus the identity $n \times n$ identity matrix corresponding to multiplication on the right of A is of the form

$$\begin{pmatrix} I_{p_1 \times p_1} & & 0 \\ & \ddots & \\ 0 & & I_{p_m \times p_m} \end{pmatrix}$$

these little identity matrices don't overlap. A similar conclusion follows from consideration of the matrices $I_{s_i \times s_i}$.

Next consider the question of multiplication of two block matrices. Let B be a block matrix of the form

$$\begin{pmatrix} B_{11} & \cdots & B_{1p} \\ \vdots & \ddots & \vdots \\ B_{r1} & \cdots & B_{rp} \end{pmatrix} \tag{11.5}$$

and A is a block matrix of the form

$$\begin{pmatrix} A_{11} & \cdots & A_{1m} \\ \vdots & \ddots & \vdots \\ A_{p1} & \cdots & A_{pm} \end{pmatrix} \tag{11.6}$$

and that for all i, j, it makes sense to multiply $B_{is}A_{sj}$ for all $s \in \{1, \cdots, p\}$. (That is the two matrices, B_{is} and A_{sj} are conformable.) and that for fixed ij, it follows $B_{is}A_{sj}$ is the same size for each s so that it makes sense to write $\sum_s B_{is}A_{sj}$.

The following theorem says essentially that when you take the product of two matrices, you can do it two ways. One way is to simply multiply them forming BA. The other way is to partition both matrices, formally multiply the blocks to get another block matrix and this one will be BA partitioned. Before presenting this theorem, here is a simple lemma which is really a special case of the theorem.

Lemma 11.2.1 *Consider the following product.*

$$\begin{pmatrix} \mathbf{0} \\ I \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & I & \mathbf{0} \end{pmatrix}$$

where the first is $n \times r$ and the second is $r \times n$. The small identity matrix I is an $r \times r$ matrix and there are l zero rows above I and l zero columns to the left of I in the right matrix. Then the product of these matrices is a block matrix of the form

$$\begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

Proof: From the definition of the way you multiply matrices, the product is

$$\begin{pmatrix} \begin{pmatrix} \mathbf{0} \\ I \\ \mathbf{0} \end{pmatrix} \mathbf{0} & \cdots \begin{pmatrix} \mathbf{0} \\ I \\ \mathbf{0} \end{pmatrix} \mathbf{0} & \begin{pmatrix} \mathbf{0} \\ I \\ \mathbf{0} \end{pmatrix} \mathbf{e}_1 & \cdots & \begin{pmatrix} \mathbf{0} \\ I \\ \mathbf{0} \end{pmatrix} \mathbf{e}_r & \begin{pmatrix} \mathbf{0} \\ I \\ \mathbf{0} \end{pmatrix} \mathbf{0} & \cdots & \begin{pmatrix} \mathbf{0} \\ I \\ \mathbf{0} \end{pmatrix} \mathbf{0} \end{pmatrix}$$

which yields the claimed result. In the formula $\mathbf{e}_j$ referrs to the column vector of length r which has a 1 in the j^{th} position. This proves the lemma.

Theorem 11.2.2 *Let B be a $q \times p$ block matrix as in 11.5 and let A be a $p \times n$ block matrix as in 11.6 such that B_{is} is conformable with A_{sj} and each product, $B_{is}A_{sj}$ for $s = 1, \cdots, p$ is of the same size so they can be added. Then BA can be obtained as a block matrix such that the ij^{th} block is of the form*

$$\sum_s B_{is}A_{sj}. \tag{11.7}$$

Proof: From 11.4

$$B_{is}A_{sj} = \begin{pmatrix} \mathbf{0} & I_{r_i \times r_i} & \mathbf{0} \end{pmatrix} B \begin{pmatrix} \mathbf{0} \\ I_{p_s \times p_s} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & I_{p_s \times p_s} & \mathbf{0} \end{pmatrix} A \begin{pmatrix} \mathbf{0} \\ I_{q_j \times q_j} \\ \mathbf{0} \end{pmatrix}$$

where here it is assumed B_{is} is $r_i \times p_s$ and A_{sj} is $p_s \times q_j$. The product involves the s^{th} block in the i^{th} row of blocks for B and the s^{th} block in the j^{th} column of A. Thus there

are the same number of rows above the $I_{p_s\times p_s}$ as there are columns to the left of $I_{p_s\times p_s}$ in those two inside matrices. Then from Lemma 11.2.1

$$\begin{pmatrix} \mathbf{0} \\ I_{p_s\times p_s} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & I_{p_s\times p_s} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_{p_s\times p_s} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

Since the blocks of small identity matrices do not overlap,

$$\sum_s \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & I_{p_s\times p_s} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} I_{p_1\times p_1} & & 0 \\ & \ddots & \\ 0 & & I_{p_p\times p_p} \end{pmatrix} = I$$

and so

$$\sum_s B_{is}A_{sj} =$$

$$\sum_s \begin{pmatrix} \mathbf{0} & I_{r_i\times r_i} & \mathbf{0} \end{pmatrix} B \begin{pmatrix} \mathbf{0} \\ I_{p_s\times p_s} \\ \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} & I_{p_s\times p_s} & \mathbf{0} \end{pmatrix} A \begin{pmatrix} \mathbf{0} \\ I_{q_j\times q_j} \\ \mathbf{0} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{0} & I_{r_i\times r_i} & \mathbf{0} \end{pmatrix} BIA \begin{pmatrix} \mathbf{0} \\ I_{q_j\times q_j} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & I_{r_i\times r_i} & \mathbf{0} \end{pmatrix} BA \begin{pmatrix} \mathbf{0} \\ I_{q_j\times q_j} \\ \mathbf{0} \end{pmatrix}$$

Hence the ij^{th} block of BA equals the formal multiplication according to matrix multiplication,

$$\sum_s B_{is}A_{sj}.$$

This proves the theorem.

Example 11.2.3 *Let an $n\times n$ matrix have the form*

$$A = \begin{pmatrix} a & \mathbf{b} \\ \mathbf{c} & P \end{pmatrix}$$

where P is $n-1\times n-1$. Multiply it by

$$B = \begin{pmatrix} p & \mathbf{q} \\ \mathbf{r} & Q \end{pmatrix}$$

where B is also an $n\times n$ matrix and Q is $n-1\times n-1$.

You use block multiplication

$$\begin{pmatrix} a & \mathbf{b} \\ \mathbf{c} & P \end{pmatrix} \begin{pmatrix} p & \mathbf{q} \\ \mathbf{r} & Q \end{pmatrix} = \begin{pmatrix} ap+\mathbf{br} & a\mathbf{q}+\mathbf{b}Q \\ p\mathbf{c}+P\mathbf{r} & \mathbf{cq}+PQ \end{pmatrix}$$

Note that this all makes sense. For example, $\mathbf{b} = 1\times n-1$ and $\mathbf{r} = n-1\times 1$ so $\mathbf{br}$ is a 1×1. Similar considerations apply to the other blocks.

Here is an interesting and significant application of block multiplication. In this theorem, $p_M(t)$ denotes the characteristic polynomial, $\det(tI-M)$. Thus the zeros of this polynomial are the eigenvalues of the matrix, M.

Theorem 11.2.4 *Let A be an $m \times n$ matrix and let B be an $n \times m$ matrix for $m \leq n$. Then*

$$p_{BA}(t) = t^{n-m} p_{AB}(t),$$

so the eigenvalues of BA and AB are the same including multiplicities except that BA has $n-m$ extra zero eigenvalues.

Proof: Use block multiplication to write

$$\begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix} \begin{pmatrix} I & A \\ 0 & I \end{pmatrix} = \begin{pmatrix} AB & ABA \\ B & BA \end{pmatrix}$$

$$\begin{pmatrix} I & A \\ 0 & I \end{pmatrix} \begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix} = \begin{pmatrix} AB & ABA \\ B & BA \end{pmatrix}.$$

Therefore,

$$\begin{pmatrix} I & A \\ 0 & I \end{pmatrix}^{-1} \begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix} \begin{pmatrix} I & A \\ 0 & I \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix}$$

Since the two matrices above are similar it follows that $\begin{pmatrix} 0 & 0 \\ B & BA \end{pmatrix}$ and $\begin{pmatrix} AB & 0 \\ B & 0 \end{pmatrix}$ have the same characteristic polynomials. Therefore, noting that BA is an $n \times n$ matrix and AB is an $m \times m$ matrix,

$$t^m \det(tI - BA) = t^n \det(tI - AB)$$

and so $\det(tI - BA) = p_{BA}(t) = t^{n-m} \det(tI - AB) = t^{n-m} p_{AB}(t)$. This proves the theorem.

11.2.2 Orthonormal Bases

Not all bases for $\mathbb{F}^n$ are created equal. Recall $\mathbb{F}$ equals either $\mathbb{C}$ or $\mathbb{R}$ and the dot product is given by

$$\mathbf{x} \cdot \mathbf{y} = \sum_j x_j \overline{y_j}.$$

The best bases are orthonormal. Much of what follows will be for $\mathbb{F}^n$ in the interest of generality.

Definition 11.2.5 *Suppose $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is a set of vectors in $\mathbb{F}^n$. It is an orthonormal set if*

$$\mathbf{v}_i \cdot \mathbf{v}_j = \delta_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

Every orthonormal set of vectors is automatically linearly independent.

Proposition 11.2.6 *Suppose $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is an orthonormal set of vectors. Then it is linearly independent.*

Proof: Suppose $\sum_{i=1}^k c_i \mathbf{v}_i = \mathbf{0}$. Then taking dot products with $\mathbf{v}_j$,

$$0 = \mathbf{0} \cdot \mathbf{v}_j = \sum_i c_i \mathbf{v}_i \cdot \mathbf{v}_j = \sum_i c_i \delta_{ij} = c_j.$$

Since j is arbitrary, this shows the set is linearly independent as claimed.

It turns out that if X is any subspace of $\mathbb{F}^m$, then there exists an orthonormal basis for X.

Lemma 11.2.7 *Let X be a subspace of $\mathbb{F}^m$ of dimension n whose basis is $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$. Then there exists an orthonormal basis for X, $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ which has the property that for each $k \leq n$, span$(\mathbf{x}_1, \cdots, \mathbf{x}_k) = \text{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k)$.*

Proof: Let $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ be a basis for X. Let $\mathbf{u}_1 \equiv \mathbf{x}_1 / |\mathbf{x}_1|$. Thus for $k = 1$, $\text{span}(\mathbf{u}_1) = \text{span}(\mathbf{x}_1)$ and $\{\mathbf{u}_1\}$ is an orthonormal set. Now suppose for some $k < n$, $\mathbf{u}_1, \cdots, \mathbf{u}_k$ have been chosen such that $(\mathbf{u}_j, \mathbf{u}_l) = \delta_{jl}$ and $\text{span}(\mathbf{x}_1, \cdots, \mathbf{x}_k) = \text{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k)$. Then define

$$\mathbf{u}_{k+1} \equiv \frac{\mathbf{x}_{k+1} - \sum_{j=1}^{k} (\mathbf{x}_{k+1} \cdot \mathbf{u}_j)\, \mathbf{u}_j}{\left| \mathbf{x}_{k+1} - \sum_{j=1}^{k} (\mathbf{x}_{k+1} \cdot \mathbf{u}_j)\, \mathbf{u}_j \right|}, \tag{11.8}$$

where the denominator is not equal to zero because the $\mathbf{x}_j$ form a basis and so

$$\mathbf{x}_{k+1} \notin \text{span}(\mathbf{x}_1, \cdots, \mathbf{x}_k) = \text{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k)$$

Thus by induction,

$$\mathbf{u}_{k+1} \in \text{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k, \mathbf{x}_{k+1}) = \text{span}(\mathbf{x}_1, \cdots, \mathbf{x}_k, \mathbf{x}_{k+1}).$$

Also, $\mathbf{x}_{k+1} \in \text{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k, \mathbf{u}_{k+1})$ which is seen easily by solving 11.8 for $\mathbf{x}_{k+1}$ and it follows

$$\text{span}(\mathbf{x}_1, \cdots, \mathbf{x}_k, \mathbf{x}_{k+1}) = \text{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k, \mathbf{u}_{k+1}).$$

If $l \leq k$,

$$\begin{aligned}
(\mathbf{u}_{k+1} \cdot \mathbf{u}_l) &= C\left((\mathbf{x}_{k+1} \cdot \mathbf{u}_l) - \sum_{j=1}^{k} (\mathbf{x}_{k+1} \cdot \mathbf{u}_j)\, (\mathbf{u}_j \cdot \mathbf{u}_l) \right) \\
&= C\left((\mathbf{x}_{k+1} \cdot \mathbf{u}_l) - \sum_{j=1}^{k} (\mathbf{x}_{k+1} \cdot \mathbf{u}_j)\, \delta_{lj} \right) \\
&= C\left((\mathbf{x}_{k+1} \cdot \mathbf{u}_l) - (\mathbf{x}_{k+1} \cdot \mathbf{u}_l) \right) = 0.
\end{aligned}$$

The vectors, $\{\mathbf{u}_j\}_{j=1}^{n}$, generated in this way are therefore an orthonormal basis because each vector has unit length.

The process by which these vectors were generated is called the Gram Schmidt process.

11.2.3 Schur's Theorem*

Every matrix is related to an upper triangular matrix in a particularly significant way. This is Shur's theorem and it is the most important theorem in the spectral theory of matrices. The important result which makes this theorem possible is the Gram Schmidt procedure of Lemma 11.2.7.

Definition 11.2.8 *An $n \times n$ matrix, U, is **unitary** if $UU^* = I = U^*U$ where U^* is defined to be the transpose of the conjugate of U. Thus $\overline{U_{ij}} = U^*_{ji}$. Note that every real orthogonal matrix is unitary. For A any matrix, A^* just defined as the conjugate of the transpose is called the **adjoint.***

Lemma 11.2.9 *The following holds. $(AB)^* = B^*A^*$.*

Proof: From the definition and remembering the properties of complex conjugation,

$$\begin{aligned}\left((AB)^*\right)_{ji} &= \overline{(AB)_{ij}} \\ &= \overline{\sum_k A_{ik}B_{kj}} = \sum_k \overline{A_{ik}B_{kj}} \\ &= \sum_k B^*_{jk}A^*_{ki} = \left(B^*A^*\right)_{ji}\end{aligned}$$

This proves the lemma.

Theorem 11.2.10 *Let A be an $n \times n$ matrix. Then there exists a unitary matrix, U such that*

$$U^*AU = T, \tag{11.9}$$

where T is an upper triangular matrix having the eigenvalues of A on the main diagonal listed according to multiplicity as roots of the characteristic equation.

Proof: Let $\mathbf{v}_1$ be a unit eigenvector for A . Then there exists λ_1 such that

$$A\mathbf{v}_1 = \lambda_1\mathbf{v}_1,\ |\mathbf{v}_1| = 1.$$

Extend $\{\mathbf{v}_1\}$ to a basis using Theorem 7.4.19 and then use the Gram Schmidt procedure to obtain $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$, an orthonormal basis in $\mathbb{F}^n$. Let U_0 be a matrix whose i^{th} column is $\mathbf{v}_i$. Then from the above, it follows U_0 is unitary. Then $U_0^*AU_0$ is of the form

$$\begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{pmatrix}$$

where A_1 is an $n-1 \times n-1$ matrix. Repeat the process for the matrix, A_1 above. There exists a unitary matrix $\widetilde{U}_1$ such that $\widetilde{U}_1^*A_1\ \widetilde{U}_1$ is of the form

$$\begin{pmatrix} \lambda_2 & * & \cdots & * \\ 0 & & & \\ \vdots & & A_2 & \\ 0 & & & \end{pmatrix}.$$

Now let U_1 be the $n \times n$ matrix of the form

$$\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \widetilde{U}_1 \end{pmatrix}.$$

This is also a unitary matrix because by block multiplication,

$$\begin{aligned}\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \widetilde{U}_1 \end{pmatrix}^* \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \widetilde{U}_1 \end{pmatrix} &= \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \widetilde{U}_1^* \end{pmatrix}\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \widetilde{U}_1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \widetilde{U}_1^*\widetilde{U}_1 \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix}\end{aligned}$$

Then using block multiplication, $U_1^*U_0^*AU_0U_1$ is of the form

$$\begin{pmatrix} \lambda_1 & * & * & \cdots & * \\ 0 & \lambda_2 & * & \cdots & * \\ 0 & 0 & & & \\ \vdots & \vdots & & A_2 & \\ 0 & 0 & & & \end{pmatrix}$$

where A_2 is an $n-2 \times n-2$ matrix. Continuing in this way, there exists a unitary matrix, U given as the product of the U_i in the above construction such that

$$U^* A U = T$$

where T is some upper triangular matrix similar to A which consequently has the same eigenvalues with the same multiplicities as A. Since the matrix is upper triangular, the characteristic equation for both A and T is $\prod_{i=1}^{n} (\lambda - \lambda_i)$ where the λ_i are the diagonal entries of T. Therefore, the λ_i are the eigenvalues.

As a simple consequence of the above theorem, here is an interesting lemma.

Lemma 11.2.11 *Let A be of the form*

$$A = \begin{pmatrix} P_1 & \cdots & * \\ \vdots & \ddots & \vdots \\ 0 & \cdots & P_s \end{pmatrix}$$

where P_k is an $m_k \times m_k$ matrix. Then

$$\det(A) = \prod_k \det(P_k).$$

Proof: Let U_k be an $m_k \times m_k$ unitary matrix such that

$$U_k^* P_k U_k = T_k$$

where T_k is upper triangular. Then letting

$$U = \begin{pmatrix} U_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & U_s \end{pmatrix},$$

it follows

$$U^* = \begin{pmatrix} U_1^* & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & U_s^* \end{pmatrix}$$

and

$$\begin{pmatrix} U_1^* & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & U_s^* \end{pmatrix} \begin{pmatrix} P_1 & \cdots & * \\ \vdots & \ddots & \vdots \\ 0 & \cdots & P_s \end{pmatrix} \begin{pmatrix} U_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & U_s \end{pmatrix}$$
$$= \begin{pmatrix} T_1 & \cdots & * \\ \vdots & \ddots & \vdots \\ 0 & \cdots & T_s \end{pmatrix}$$

and so

$$\det(A) = \prod_k \det(T_k) = \prod_k \det(P_k).$$

This proves the lemma.

Definition 11.2.12 *An $n \times n$ matrix, A is called **Hermitian** if $A = A^*$. Thus a real symmetric matrix is Hermitian.*

Theorem 11.2.13 *If A is Hermitian, there exists a unitary matrix, U such that*

$$U^*AU = D \tag{11.10}$$

where D is a diagonal matrix. That is, D has nonzero entries only on the main diagonal. Furthermore, the columns of U are an orthonormal basis for $\mathbb{F}^n$.

Proof: From Schur's theorem above, there exists U unitary such that

$$U^*AU = T$$

where T is an upper triangular matrix. Then from Lemma 11.2.9

$$T^* = (U^*AU)^* = U^*A^*U = T.$$

Thus $T = T^*$ and T is upper triangular. This can only happen if T is really a diagonal matrix. (If $i \neq j$, one of T_{ij} or T_{ji} equals zero. But $T_{ij} = \overline{T_{ji}}$ and so they are both zero.

Finally, let

$$U = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$$

where the $\mathbf{u}_i$ denote the columns of U and

$$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

The equation, $U^*AU = D$ implies

$$\begin{aligned} AU &= \begin{pmatrix} A\mathbf{u}_1 & A\mathbf{u}_2 & \cdots & A\mathbf{u}_n \end{pmatrix} \\ &= UD = \begin{pmatrix} \lambda_1\mathbf{u}_1 & \lambda_2\mathbf{u}_2 & \cdots & \lambda_n\mathbf{u}_n \end{pmatrix} \end{aligned}$$

where the entries denote the columns of AU and UD respectively. Therefore, $A\mathbf{u}_i = \lambda_i\mathbf{u}_i$ and since the matrix is unitary, the ij^{th} entry of U^*U equals δ_{ij} and so

$$\delta_{ij} = \overline{\mathbf{u}}_i^T\mathbf{u}_j = \overline{\mathbf{u}_i^T\overline{\mathbf{u}}_j} = \overline{\mathbf{u}_i \cdot \overline{\mathbf{u}_j}}.$$

This proves the corollary because it shows the vectors $\{\mathbf{u}_i\}$ form an orthonormal basis. This proves the theorem.

Corollary 11.2.14 *If A is Hermitian, then all the eigenvalues of A are real.*

Proof: Since A is Hermitian, there exists unitary, U such that $U^*AU = D$, a diagonal matrix whose diagonal entries are the eigenvalues of A. Therefore, $D^* = U^*A^*U = U^*AU = D$ showing D is real.

Corollary 11.2.15 *If A is a real symmetric ($A = A^T$)matrix, then A is Hermitian and there exists a real unitary matrix, U such that $U^TAU = D$ where D is a diagonal matrix.*

Proof: This follows from Corollary 11.2.14 which says the eigenvalues are all real. Then if $A\mathbf{x} = \lambda\mathbf{x}$, the same is true of $\overline{\mathbf{x}}$. and so in the construction for Shur's theorem, you can always deal exclusively with real eigenvectors as long as your matrices are real and symmetric. When you construct the matrix which reduces the problem to a smaller one having A_1 in the lower right corner, use the Gram Schmidt process on $\mathbb{R}^n$ using the real dot product to construct vectors, $\mathbf{v}_2, \cdots, \mathbf{v}_n$ in $\mathbb{R}^n$ such that $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is an orthonormal basis for $\mathbb{R}^n$. The matrix A_1 is symmetric also. This is because for $j, k \geq 2$

$$A_{1kj} = \mathbf{v}_k^TA\mathbf{v}_j = \left(\mathbf{v}_k^TA\mathbf{v}_j\right)^T = \mathbf{v}_j^TA\mathbf{v}_k = A_{1jk}.$$

Therefore, continuing this way, the process of the proof delivers only real vectors and real matrices.

11.3 Least Square Approximation

A very important technique is that of the least square approximation.

Lemma 11.3.1 *Let A be an $m \times n$ matrix and let $A(\mathbb{F}^n)$ denote the set of vectors in $\mathbb{F}^m$ which are of the form $A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{F}^n$. Then $A(\mathbb{F}^n)$ is a subspace of $\mathbb{F}^m$.*

Proof: Let $A\mathbf{x}$ and $A\mathbf{y}$ be two points of $A(\mathbb{F}^n)$. It suffices to verify that if a, b are scalars, then $aA\mathbf{x} + bA\mathbf{y}$ is also in $A(\mathbb{F}^n)$. But $aA\mathbf{x} + bA\mathbf{y} = A(a\mathbf{x} + b\mathbf{y})$ because A is linear. This proves the lemma.

Theorem 11.3.2 *Let $\mathbf{y} \in \mathbb{F}^m$ and let A be an $m \times n$ matrix. Then there exists $\mathbf{x} \in \mathbb{F}^n$ minimizing the function, $|\mathbf{y}-A\mathbf{x}|^2$. Furthermore, $\mathbf{x}$ minimizes this function if and only if*

$$(\mathbf{y}-A\mathbf{x}) \cdot A\mathbf{w} = 0$$

for all $\mathbf{w} \in \mathbb{F}^n$.

Proof: Let $\{\mathbf{f}_1, \cdots, \mathbf{f}_r\}$ be an orthonormal basis for $A(\mathbb{F}^n)$. Since

$$A(\mathbb{F}^n) = \operatorname{span}(\mathbf{f}_1, \cdots, \mathbf{f}_r),$$

it follows that there exists $y_1, \cdots, y_r$ that minimize

$$\left| \mathbf{y} - \sum_{k=1}^{r} y_k \mathbf{f}_k \right|^2,$$

then letting $A\mathbf{x} = \sum_{k=1}^{r} y_k \mathbf{f}_k$, it will follow that this $\mathbf{x}$ is the desired solution. Now here are the details.

Let $y_1, \cdots, y_r$ be a list of scalars in $\mathbb{F}$. Then from the definition of $|\cdot|$ and the properties of the dot product,

$$\begin{aligned}
\left| \mathbf{y} - \sum_{k=1}^{r} y_k \mathbf{f}_k \right|^2 &= \left(\mathbf{y} - \sum_{k=1}^{r} y_k \mathbf{f}_k \right) \cdot \left(\mathbf{y} - \sum_{k=1}^{r} y_k \mathbf{f}_k \right) \\
&= |\mathbf{y}|^2 - 2 \operatorname{Re} \sum_{k=1}^{r} y_k (\mathbf{y} \cdot \mathbf{f}_k) + \sum_k \sum_l y_k \overline{y_l} \overbrace{\mathbf{f}_k \cdot \mathbf{f}_l}^{\delta_{kl}} \\
&= |\mathbf{y}|^2 - 2 \operatorname{Re} \sum_{k=1}^{r} y_k (\mathbf{y} \cdot \mathbf{f}_k) + \sum_{k=1}^{r} |y_k|^2 \\
&= |\mathbf{y}|^2 + \sum_{k=1}^{r} |y_k|^2 - 2 \operatorname{Re} y_k (\mathbf{y} \cdot \mathbf{f}_k)
\end{aligned}$$

Now complete the square to obtain

$$\begin{aligned}
&= |\mathbf{y}|^2 + \sum_{k=1}^{r} \left(|y_k|^2 - 2 \operatorname{Re} y_k (\mathbf{y} \cdot \mathbf{f}_k) + |\mathbf{y} \cdot \mathbf{f}_k|^2 \right) - \sum_{k=1}^{r} |\mathbf{y} \cdot \mathbf{f}_k|^2 \\
&= |\mathbf{y}|^2 + \sum_{k=1}^{r} |y_k - (\mathbf{y} \cdot \mathbf{f}_k)|^2 - \sum_{k=1}^{r} (\mathbf{y} \cdot \mathbf{f}_k)^2 .
\end{aligned}$$

This shows that the minimum is obtained when $y_k = (\mathbf{y} \cdot \mathbf{f}_k)$. This proves the existence part of the Theorem.

To verify the other part, let $t \in \mathbb{R}$ and consider

$$\begin{aligned} |\mathbf{y}-A(\mathbf{x}+t\mathbf{w})|^2 &= (\mathbf{y}-A\mathbf{x}-tA\mathbf{w})\cdot(\mathbf{y}-A\mathbf{x}-tA\mathbf{w}) \\ &= |\mathbf{y}-A\mathbf{x}|^2 - 2t\,\mathrm{Re}\,(\mathbf{y}-A\mathbf{x})\cdot A\mathbf{w} + t^2|A\mathbf{w}|^2. \end{aligned}$$

Then from the above equation, $|\mathbf{y}-A\mathbf{x}|^2 \leq |\mathbf{y}-A\mathbf{z}|^2$ for all $\mathbf{z} \in \mathbb{F}^n$ if and only if for all $\mathbf{w} \in \mathbb{F}^n$ and $t \in \mathbb{R}$

$$|\mathbf{y}-A\mathbf{x}|^2 - 2t\,\mathrm{Re}\,(\mathbf{y}-A\mathbf{x})\cdot A\mathbf{w} + t^2|\mathbf{w}|^2 \geq |\mathbf{y}-A\mathbf{x}|^2$$

and this happens if and only if for all $t \in \mathbb{R}$ and $\mathbf{w} \in \mathbb{F}^n$,

$$-2t\,\mathrm{Re}\,(\mathbf{y}-A\mathbf{x})\cdot A\mathbf{w} + t^2|A\mathbf{w}|^2 \geq 0,$$

which occurs if and only if $\mathrm{Re}\,(\mathbf{y}-A\mathbf{x})\cdot A\mathbf{w} = 0$ for all $\mathbf{w} \in \mathbb{R}^n$. (Why?)

This implies that $(\mathbf{y}-A\mathbf{x})\cdot A\mathbf{w} = 0$ for every $\mathbf{w} \in \mathbb{F}^n$ because there exists a complex number, θ of magnitude 1 such that

$$\begin{aligned} |(\mathbf{y}-A\mathbf{x})\cdot A\mathbf{w}| &= \theta(\mathbf{y}-A\mathbf{x})\cdot A\mathbf{w} = (\mathbf{y}-A\mathbf{x})\cdot A\overline{\theta}\mathbf{w} \\ &= \mathrm{Re}\,(\mathbf{y}-A\mathbf{x})\cdot A\overline{\theta}\mathbf{w} = 0. \end{aligned}$$

This proves the theorem.

Recall the definition of the adjoint of a matrix.

Definition 11.3.3 *Let A be an $m \times n$ matrix. Then*

$$A^* \equiv \overline{(A^T)}.$$

This means you take the transpose of A and then replace each entry by its conjugate. This matrix is called the ***adjoint****. Thus in the case of real matrices having only real entries, the adjoint is just the transpose.*

Lemma 11.3.4 *Let A be an $m \times n$ matrix. Then*

$$A\mathbf{x}\cdot\mathbf{y} = \mathbf{x}\cdot A^*\mathbf{y}$$

Proof: This follows from the definition.

$$\begin{aligned} A\mathbf{x}\cdot\mathbf{y} &= \sum_{i,j} A_{ij}x_j\overline{y_i} \\ &= \sum_{i,j} x_j\overline{A^*_{ji}}\,\overline{y_i} \\ &= \mathbf{x}\cdot A^*\mathbf{y}. \end{aligned}$$

This proves the lemma.

The next corollary gives the technique of least squares.

Corollary 11.3.5 *A value of $\mathbf{x}$ which solves the problem of Theorem 11.3.2 is obtained by solving the equation*

$$A^*A\mathbf{x} = A^*\mathbf{y}$$

and furthermore, there exists a solution to this system of equations.

Proof: For $\mathbf{x}$ the unique minimizer of Theorem 11.3.2, $(\mathbf{y}-A\mathbf{x})\cdot A\mathbf{w} = 0$ for all $\mathbf{w} \in \mathbb{F}^n$ and from Lemma 11.3.4, this is the same as saying

$$A^*(\mathbf{y}-A\mathbf{x})\cdot\mathbf{w} = 0$$

for all $\mathbf{w} \in \mathbb{F}^n$. This implies

$$A^*\mathbf{y} - A^*A\mathbf{x} = \mathbf{0}.$$

Therefore, there is a unique solution to the equation of this corollary and it solves the minimization problem of Theorem 11.3.2.

11.3.1 The Least Squares Regression Line

For the situation of the least squares regression line discussed here I will specialize to the case of $\mathbb{R}^n$ rather than $\mathbb{F}^n$ because it seems this case is by far the most interesting and the extra details are not justified by an increase in utility. Thus, everywhere you see A^* it suffices to place A^T.

An important application of Corollary 11.3.5 is the problem of finding the least squares regression line in statistics. Suppose you are given points in the plane, $\{(x_i, y_i)\}_{i=1}^n$ and you would like to find constants m and b such that the line $y = mx + b$ goes through all these points. Of course this will be impossible in general. Therefore, try to find m, b to get as close as possible. The desired system is

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix}$$

which is of the form $\mathbf{y} = A\mathbf{x}$ and it is desired to choose m and b to make

$$\left| A \begin{pmatrix} m \\ b \end{pmatrix} - \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right|^2$$

as small as possible. According to Theorem 11.3.2 and Corollary 11.3.5, the best values for m and b occur as the solution to

$$A^T A \begin{pmatrix} m \\ b \end{pmatrix} = A^T \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

where

$$A = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}.$$

Thus, computing $A^T A$,

$$\begin{pmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n y_i \end{pmatrix}$$

Solving this system of equations for m and b,

$$m = \frac{-\left(\sum_{i=1}^n x_i\right)\left(\sum_{i=1}^n y_i\right) + \left(\sum_{i=1}^n x_i y_i\right) n}{\left(\sum_{i=1}^n x_i^2\right) n - \left(\sum_{i=1}^n x_i\right)^2}$$

and

$$b = \frac{-\left(\sum_{i=1}^n x_i\right)\sum_{i=1}^n x_i y_i + \left(\sum_{i=1}^n y_i\right)\sum_{i=1}^n x_i^2}{\left(\sum_{i=1}^n x_i^2\right) n - \left(\sum_{i=1}^n x_i\right)^2}.$$

One could clearly do a least squares fit for curves of the form $y = ax^2 + bx + c$ in the same way. In this case you want to solve as well as possible for a, b, and c the system

$$\begin{pmatrix} x_1^2 & x_1 & 1 \\ \vdots & \vdots & \vdots \\ x_n^2 & x_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

and one would use the same technique as above. Many other similar problems are important, including many in higher dimensions and they are all solved the same way.

11.3.2 The Fredholm Alternative

The next major result is called the Fredholm alternative. It comes from Theorem 11.3.2 and Lemma 11.3.4.

Theorem 11.3.6 *Let A be an $m \times n$ matrix. Then there exists $\mathbf{x} \in \mathbb{F}^n$ such that $A\mathbf{x} = \mathbf{y}$ if and only if whenever $A^*\mathbf{z} = \mathbf{0}$ it follows that $\mathbf{z} \cdot \mathbf{y} = 0$.*

Proof: First suppose that for some $\mathbf{x} \in \mathbb{F}^n$, $A\mathbf{x} = \mathbf{y}$. Then letting $A^*\mathbf{z} = \mathbf{0}$ and using Lemma 11.3.4

$$\mathbf{y} \cdot \mathbf{z} = A\mathbf{x} \cdot \mathbf{z} = \mathbf{x} \cdot A^*\mathbf{z} = \mathbf{x} \cdot \mathbf{0} = 0.$$

This proves half the theorem.

To do the other half, suppose that whenever, $A^*\mathbf{z} = \mathbf{0}$ it follows that $\mathbf{z} \cdot \mathbf{y} = 0$. It is necessary to show there exists $\mathbf{x} \in \mathbb{F}^n$ such that $\mathbf{y} = A\mathbf{x}$. From Theorem 11.3.2 there exists $\mathbf{x}$ minimizing $|\mathbf{y} - A\mathbf{x}|^2$ which therefore satisfies

$$(\mathbf{y} - A\mathbf{x}) \cdot A\mathbf{w} = 0 \tag{11.11}$$

for all $\mathbf{w} \in \mathbb{F}^n$. Therefore, for all $\mathbf{w} \in \mathbb{F}^n$,

$$A^* (\mathbf{y} - A\mathbf{x}) \cdot \mathbf{w} = 0$$

which shows that $A^* (\mathbf{y} - A\mathbf{x}) = \mathbf{0}$. (Why?) Therefore, by assumption,

$$(\mathbf{y} - A\mathbf{x}) \cdot \mathbf{y} = 0.$$

Now by 11.11 with $\mathbf{w} = \mathbf{x}$,

$$(\mathbf{y} - A\mathbf{x}) \cdot (\mathbf{y} - A\mathbf{x}) = (\mathbf{y} - A\mathbf{x}) \cdot \mathbf{y} - (\mathbf{y} - A\mathbf{x}) \cdot A\mathbf{x} = 0$$

showing that $\mathbf{y} = A\mathbf{x}$. This proves the theorem.

The following corollary is also called the Fredholm alternative.

Corollary 11.3.7 *Let A be an $m \times n$ matrix. Then A is onto if and only if A^* is one to one.*

Proof: Suppose first A is onto. Then by Theorem 11.3.6, it follows that for all $\mathbf{y} \in \mathbb{F}^m$, $\mathbf{y} \cdot \mathbf{z} = 0$ whenever $A^*\mathbf{z} = \mathbf{0}$. Therefore, let $\mathbf{y} = \mathbf{z}$ where $A^*\mathbf{z} = \mathbf{0}$ and conclude that $\mathbf{z} \cdot \mathbf{z} = 0$ whenever $A^*\mathbf{z} = 0$. If $A^*\mathbf{x} = A^*\mathbf{y}$, then $A^* (\mathbf{x} - \mathbf{y}) = \mathbf{0}$ and so $\mathbf{x} - \mathbf{y} = \mathbf{0}$. Thus A^* is one to one.

Now let $\mathbf{y} \in \mathbb{F}^m$ be given. $\mathbf{y} \cdot \mathbf{z} = 0$ whenever $A^*\mathbf{z} = \mathbf{0}$ because, since A^* is assumed to be one to one, and $\mathbf{0}$ is a solution to this equation, it must be the only solution. Therefore, by Theorem 11.3.6 there exists $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{y}$ therefore, A is onto.

11.4 The Right Polar Factorization*

The right polar factorization involves writing a matrix as a product of two other matrices, one which preserves distances and the other which stretches and distorts. First here are some lemmas which review and add to many of the topics discussed so far about adjoints and orthonormal sets and such things.

Lemma 11.4.1 *Let A be a Hermitian matrix such that all its eigenvalues are nonnegative. Then there exists a Hermitian matrix, $A^{1/2}$ such that $A^{1/2}$ has all nonnegative eigenvalues and $\left(A^{1/2}\right)^2 = A$.*

Proof: Since A is Hermitian, there exists a diagonal matrix D having all real nonnegative entries and a unitary matrix U such that $A = U^*DU$. Then denote by $D^{1/2}$ the matrix which is obtained by replacing each diagonal entry of D with its square root. Thus $D^{1/2}D^{1/2} = D$. Then define

$$A^{1/2} \equiv U^*D^{1/2}U.$$

Then

$$\left(A^{1/2}\right)^2 = U^*D^{1/2}UU^*D^{1/2}U = U^*DU = A.$$

Since $D^{1/2}$ is real,

$$\left(U^*D^{1/2}U\right)^* = U^*\left(D^{1/2}\right)^*(U^*)^* = U^*D^{1/2}U$$

so $A^{1/2}$ is Hermitian. This proves the lemma.

There is also a useful observation about orthonormal sets of vectors which is stated in the next lemma.

Lemma 11.4.2 *Suppose $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_r\}$ is an orthonormal set of vectors. Then if $c_1, \cdots, c_r$ are scalars,*

$$\left|\sum_{k=1}^{r} c_k\mathbf{x}_k\right|^2 = \sum_{k=1}^{r} |c_k|^2 .$$

Proof: This follows from the definition. From the properties of the dot product and using the fact that the given set of vectors is orthonormal,

$$\left|\sum_{k=1}^{r} c_k\mathbf{x}_k\right|^2 = \left(\sum_{k=1}^{r} c_k\mathbf{x}_k, \sum_{j=1}^{r} c_j\mathbf{x}_j\right)$$

$$= \sum_{k,j} c_k\overline{c_j}\left(\mathbf{x}_k, \mathbf{x}_j\right) = \sum_{k=1}^{r} |c_k|^2 .$$

This proves the lemma.

Next it is helpful to recall the Gram Schmidt algorithm and observe a certain property stated in the next lemma.

Lemma 11.4.3 *Suppose $\{\mathbf{w}_1, \cdots, \mathbf{w}_r, \mathbf{v}_{r+1}, \cdots, \mathbf{v}_p\}$ is a linearly independent set of vectors such that $\{\mathbf{w}_1, \cdots, \mathbf{w}_r\}$ is an orthonormal set of vectors. Then when the Gram Schmidt process is applied to the vectors in the given order, it will not change any of the $\mathbf{w}_1, \cdots, \mathbf{w}_r$.*

Proof: Let $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$ be the orthonormal set delivered by the Gram Schmidt process. Then $\mathbf{u}_1 = \mathbf{w}_1$ because by definition, $\mathbf{u}_1 \equiv \mathbf{w}_1/|\mathbf{w}_1| = \mathbf{w}_1$. Now suppose $\mathbf{u}_j = \mathbf{w}_j$ for all $j \leq k \leq r$. Then if $k < r$, consider the definition of $\mathbf{u}_{k+1}$.

$$\mathbf{u}_{k+1} \equiv \frac{\mathbf{w}_{k+1} - \sum_{j=1}^{k+1}\left(\mathbf{w}_{k+1}, \mathbf{u}_j\right)\mathbf{u}_j}{\left|\mathbf{w}_{k+1} - \sum_{j=1}^{k+1}\left(\mathbf{w}_{k+1}, \mathbf{u}_j\right)\mathbf{u}_j\right|}$$

By induction, $\mathbf{u}_j = \mathbf{w}_j$ and so this reduces to $\mathbf{w}_{k+1}/|\mathbf{w}_{k+1}| = \mathbf{w}_{k+1}$. This proves the lemma.

This lemma immediately implies the following lemma.

Lemma 11.4.4 *Let V be a subspace of dimension p and let $\{\mathbf{w}_1, \cdots, \mathbf{w}_r\}$ be an orthonormal set of vectors in V. Then this orthonormal set of vectors may be extended to an orthonormal basis for V,*

$$\{\mathbf{w}_1, \cdots, \mathbf{w}_r, \mathbf{y}_{r+1}, \cdots, \mathbf{y}_p\}$$

Proof: First extend the given linearly independent set $\{\mathbf{w}_1, \cdots, \mathbf{w}_r\}$ to a basis for V and then apply the Gram Schmidt theorem to the resulting basis. Since $\{\mathbf{w}_1, \cdots, \mathbf{w}_r\}$ is orthonormal it follows from Lemma 11.4.3 the result is of the desired form, an orthonormal basis extending $\{\mathbf{w}_1, \cdots, \mathbf{w}_r\}$. This proves the lemma.

Here is another lemma about preserving distance.

Lemma 11.4.5 *Suppose R is an $m \times n$ matrix with $m > n$ and R preserves distances. Then $R^*R = I$.*

Proof: Since R preserves distances, $|R\mathbf{x}| = |\mathbf{x}|$ for every $\mathbf{x}$. Therefore from the axioms of the dot product,

$$\begin{aligned} & |\mathbf{x}|^2 + |\mathbf{y}|^2 + (\mathbf{x}, \mathbf{y}) + (\mathbf{y}, \mathbf{x}) \\ = \; & |\mathbf{x} + \mathbf{y}|^2 \\ = \; & (R(\mathbf{x} + \mathbf{y}), R(\mathbf{x} + \mathbf{y})) \\ = \; & (R\mathbf{x}, R\mathbf{x}) + (R\mathbf{y}, R\mathbf{y}) + (R\mathbf{x}, R\mathbf{y}) + (R\mathbf{y}, R\mathbf{x}) \\ = \; & |\mathbf{x}|^2 + |\mathbf{y}|^2 + (R^*R\mathbf{x}, \mathbf{y}) + (\mathbf{y}, R^*R\mathbf{x}) \end{aligned}$$

and so for all $\mathbf{x}, \mathbf{y}$,

$$(R^*R\mathbf{x} - \mathbf{x}, \mathbf{y}) + (\mathbf{y}, R^*R\mathbf{x} - \mathbf{x}) = 0$$

Hence for all $\mathbf{x}, \mathbf{y}$,

$$\mathrm{Re}\,(R^*R\mathbf{x} - \mathbf{x}, \mathbf{y}) = 0$$

Now for a $\mathbf{x}, \mathbf{y}$ given, choose $\alpha \in \mathbb{C}$ such that

$$\alpha\,(R^*R\mathbf{x} - \mathbf{x}, \mathbf{y}) = |(R^*R\mathbf{x} - \mathbf{x}, \mathbf{y})|$$

Then

$$\begin{aligned} 0 \; &= \; \mathrm{Re}\,(R^*R\mathbf{x} - \mathbf{x}, \overline{\alpha}\mathbf{y}) = \mathrm{Re}\,\alpha\,(R^*R\mathbf{x} - \mathbf{x}, \mathbf{y}) \\ &= \; |(R^*R\mathbf{x} - \mathbf{x}, \mathbf{y})| \end{aligned}$$

Thus $|(R^*R\mathbf{x} - \mathbf{x}, \mathbf{y})| = 0$ for all $\mathbf{x}, \mathbf{y}$ because the given $\mathbf{x}, \mathbf{y}$ were arbitrary. Let $\mathbf{y} = R^*R\mathbf{x} - \mathbf{x}$ to conclude that for all $\mathbf{x}$,

$$R^*R\mathbf{x} - \mathbf{x} = \mathbf{0}$$

which says $R^*R = I$ since $\mathbf{x}$ is arbitrary. This proves the lemma.

With this preparation, here is the big theorem about the right polar factorization.

Theorem 11.4.6 *Let F be an $m \times n$ matrix where $m \geq n$. Then there exists a Hermitian $n \times n$ matrix, U which has all nonnegative eigenvalues and an $m \times n$ matrix, R which preserves distances and satisfies $R^*R = I$ such that*

$$F = RU.$$

Proof: Consider F^*F. This is a Hermitian matrix because

$$(F^*F)^* = F^*(F^*)^* = F^*F$$

Also the eigenvalues of the $n \times n$ matrix F^*F are all nonnegative. This is because if $\mathbf{x}$ is an eigenvalue,

$$\lambda(\mathbf{x},\mathbf{x}) = (F^*F\mathbf{x},\mathbf{x}) = (F\mathbf{x},F\mathbf{x}) \geq 0.$$

Therefore, by Lemma 11.4.1, there exists an $n \times n$ Hermitian matrix, U having all nonnegative eigenvalues such that

$$U^2 = F^*F.$$

Consider the subspace $U(\mathbb{F}^n)$. Let $\{U\mathbf{x}_1, \cdots, U\mathbf{x}_r\}$ be an orthonormal basis for $U(\mathbb{F}^n) \subseteq \mathbb{F}^n$. Note that $U(\mathbb{F}^n)$ might not be all of $\mathbb{F}^n$. Using Lemma 11.4.4, extend to an orthonormal basis for all of $\mathbb{F}^n$,

$$\{U\mathbf{x}_1, \cdots, U\mathbf{x}_r, \mathbf{y}_{r+1}, \cdots, \mathbf{y}_n\}.$$

Next observe that $\{F\mathbf{x}_1, \cdots, F\mathbf{x}_r\}$ is also an orthonormal set of vectors in $\mathbb{F}^m$. This is because

$$\begin{aligned}(F\mathbf{x}_k, F\mathbf{x}_j) &= (F^*F\mathbf{x}_k, \mathbf{x}_j) = \left(U^2\mathbf{x}_k, \mathbf{x}_j\right) \\ &= (U\mathbf{x}_k, U^*\mathbf{x}_j) = (U\mathbf{x}_k, U\mathbf{x}_j) = \delta_{jk}\end{aligned}$$

Therefore, from Lemma 11.4.4 again, this orthonormal set of vectors can be extended to an orthonormal basis for $\mathbb{F}^m$,

$$\{F\mathbf{x}_1, \cdots, F\mathbf{x}_r, \mathbf{z}_{r+1}, \cdots, \mathbf{z}_m\}$$

Thus there are at least as many $\mathbf{z}_k$ as there are $\mathbf{y}_j$. Now for $\mathbf{x} \in \mathbb{F}^n$, since

$$\{U\mathbf{x}_1, \cdots, U\mathbf{x}_r, \mathbf{y}_{r+1}, \cdots, \mathbf{y}_n\}$$

is an orthonormal basis for $\mathbb{F}^n$, there exist unique scalars,

$$c_1 \cdots, c_r, d_{r+1}, \cdots, d_n$$

such that

$$\mathbf{x} = \sum_{k=1}^{r} c_k U\mathbf{x}_k + \sum_{j=r+1}^{n} d_k \mathbf{y}_k$$

Define

$$R\mathbf{x} \equiv \sum_{k=1}^{r} c_k F\mathbf{x}_k + \sum_{j=r+1}^{n} d_k \mathbf{z}_k \tag{11.12}$$

Then also there exist scalars b_k such that

$$U\mathbf{x} = \sum_{k=1}^{r} b_k U\mathbf{x}_k$$

and so from 11.12,

$$RU\mathbf{x} = \sum_{k=1}^{r} b_k F\mathbf{x}_k = F\left(\sum_{k=1}^{r} b_k \mathbf{x}_k\right)$$

Is $F\left(\sum_{k=1}^{r} b_k \mathbf{x}_k\right) = F(\mathbf{x})$?

$$\left(F\left(\sum_{k=1}^{r} b_k \mathbf{x}_k\right) - F(\mathbf{x}), F\left(\sum_{k=1}^{r} b_k \mathbf{x}_k\right) - F(\mathbf{x})\right)$$

$$\begin{aligned}
&= \left((F^*F) \left(\sum_{k=1}^r b_k \mathbf{x}_k - \mathbf{x} \right), \left(\sum_{k=1}^r b_k \mathbf{x}_k - \mathbf{x} \right) \right) \\
&= \left(U^2 \left(\sum_{k=1}^r b_k \mathbf{x}_k - \mathbf{x} \right), \left(\sum_{k=1}^r b_k \mathbf{x}_k - \mathbf{x} \right) \right) \\
&= \left(U \left(\sum_{k=1}^r b_k \mathbf{x}_k - \mathbf{x} \right), U \left(\sum_{k=1}^r b_k \mathbf{x}_k - \mathbf{x} \right) \right) \\
&= \left(\sum_{k=1}^r b_k U\mathbf{x}_k - U\mathbf{x}, \sum_{k=1}^r b_k U\mathbf{x}_k - U\mathbf{x} \right) = 0
\end{aligned}$$

Therefore, $F\left(\sum_{k=1}^r b_k \mathbf{x}_k\right) = F(\mathbf{x})$ and this shows

$$RU\mathbf{x} = F\mathbf{x}.$$

From 11.12 and Lemma 11.4.2 R preserves distances. Therefore, by Lemma 11.4.5 $R^*R = I$. This proves the theorem.

11.5 The Singular Value Decomposition*

In this section, A will be an $m \times n$ matrix. To begin with, here is a simple lemma.

Lemma 11.5.1 *Let A be an $m \times n$ matrix. Then A^*A is self adjoint and all its eigenvalues are nonnegative.*

Proof: It is obvious that A^*A is self adjoint. Suppose $A^*A\mathbf{x} = \lambda\mathbf{x}$. Then $\lambda |\mathbf{x}|^2 = (\lambda\mathbf{x}, \mathbf{x}) = (A^*A\mathbf{x}, \mathbf{x}) = (A\mathbf{x}, A\mathbf{x}) \geq 0$.

Definition 11.5.2 *Let A be an $m \times n$ matrix. The singular values of A are the square roots of the positive eigenvalues of A^*A.*

With this definition and lemma here is the main theorem on the singular value decomposition.

Theorem 11.5.3 *Let A be an $m \times n$ matrix. Then there exist unitary matrices, U and V of the appropriate size such that*

$$U^*AV = \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix}$$

where σ is of the form

$$\sigma = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_k \end{pmatrix}$$

for the σ_i the singular values of A.

Proof: By the above lemma and Theorem 11.2.13 there exists an orthonormal basis, $\{\mathbf{v}_i\}_{i=1}^n$ such that $A^*A\mathbf{v}_i = \sigma_i^2\mathbf{v}_i$ where $\sigma_i^2 > 0$ for $i = 1, \cdots, k, (\sigma_i > 0)$ and equals zero if $i > k$. Thus for $i > k$, $A\mathbf{v}_i = \mathbf{0}$ because

$$(A\mathbf{v}_i, A\mathbf{v}_i) = (A^*A\mathbf{v}_i, \mathbf{v}_i) = (\mathbf{0}, \mathbf{v}_i) = 0.$$

For $i = 1, \cdots, k$, define $\mathbf{u}_i \in \mathbb{F}^m$ by

$$\mathbf{u}_i \equiv \sigma_i^{-1} A\mathbf{v}_i.$$

Thus $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$. Now

$$\begin{aligned}(\mathbf{u}_i, \mathbf{u}_j) &= \left(\sigma_i^{-1} A\mathbf{v}_i, \sigma_j^{-1} A\mathbf{v}_j\right) = \left(\sigma_i^{-1}\mathbf{v}_i, \sigma_j^{-1} A^* A\mathbf{v}_j\right) \\ &= \left(\sigma_i^{-1}\mathbf{v}_i, \sigma_j^{-1}\sigma_j^2 \mathbf{v}_j\right) = \frac{\sigma_j}{\sigma_i}(\mathbf{v}_i, \mathbf{v}_j) = \delta_{ij}.\end{aligned}$$

Thus $\{\mathbf{u}_i\}_{i=1}^k$ is an orthonormal set of vectors in $\mathbb{F}^m$. Also,

$$AA^*\mathbf{u}_i = AA^*\sigma_i^{-1} A\mathbf{v}_i = \sigma_i^{-1} AA^* A\mathbf{v}_i = \sigma_i^{-1} A\sigma_i^2 \mathbf{v}_i = \sigma_i^2 \mathbf{u}_i.$$

Now extend $\{\mathbf{u}_i\}_{i=1}^k$ to an orthonormal basis for all of $\mathbb{F}^m$, $\{\mathbf{u}_i\}_{i=1}^m$ and let $U \equiv (\mathbf{u}_1 \cdots \mathbf{u}_m)$ while $V \equiv (\mathbf{v}_1 \cdots \mathbf{v}_n)$. Thus U is the matrix which has the $\mathbf{u}_i$ as columns and V is defined as the matrix which has the $\mathbf{v}_i$ as columns. Then

$$\begin{aligned}U^*AV &= \begin{pmatrix} \mathbf{u}_1^* \\ \vdots \\ \mathbf{u}_k^* \\ \vdots \\ \mathbf{u}_m^* \end{pmatrix} A(\mathbf{v}_1 \cdots \mathbf{v}_n) \\ &= \begin{pmatrix} \mathbf{u}_1^* \\ \vdots \\ \mathbf{u}_k^* \\ \vdots \\ \mathbf{u}_m^* \end{pmatrix} (\sigma_1 \mathbf{u}_1 \cdots \sigma_k \mathbf{u}_k \mathbf{0} \cdots \mathbf{0}) \\ &= \begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix}\end{aligned}$$

where σ is given in the statement of the theorem.

The singular value decomposition has as an immediate corollary the following interesting result.

Corollary 11.5.4 *Let A be an $m \times n$ matrix. Then the rank of A and A^* equals the number of singular values.*

Proof: Since V and U are unitary, it follows that

$$\begin{aligned}\operatorname{rank}(A) &= \operatorname{rank}(U^*AV) \\ &= \operatorname{rank}\begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix} \\ &= \text{number of singular values.}\end{aligned}$$

Also since U, V are unitary,

$$\begin{aligned}\operatorname{rank}(A^*) &= \operatorname{rank}(V^*A^*U) \\ &= \operatorname{rank}\left((U^*AV)^*\right) \\ &= \operatorname{rank}\left(\begin{pmatrix} \sigma & 0 \\ 0 & 0 \end{pmatrix}^*\right) \\ &= \text{number of singular values.}\end{aligned}$$

This proves the corollary.

The singular value decomposition also has a very interesting connection to the problem of least squares solutions. Recall that it was desired to find $\mathbf{x}$ such that $|A\mathbf{x}-\mathbf{y}|$ is as small as possible. Lemma 11.3.2 shows that there is a solution to this problem which can be found by solving the system $A^*A\mathbf{x} = A^*\mathbf{y}$. Each $\mathbf{x}$ which solves this system solves the minimization problem as was shown in the lemma just mentioned. Now consider this equation for the solutions of the minimization problem in terms of the singular value decomposition.

$$\overbrace{V\begin{pmatrix}\sigma & 0\\ 0 & 0\end{pmatrix}U^*}^{A^*}\overbrace{U\begin{pmatrix}\sigma & 0\\ 0 & 0\end{pmatrix}V^*}^{A}\mathbf{x} = \overbrace{V\begin{pmatrix}\sigma & 0\\ 0 & 0\end{pmatrix}U^*}^{A^*}\mathbf{y}.$$

Therefore, this yields the following upon using block multiplication and multiplying on the left by V^*.

$$\begin{pmatrix}\sigma^2 & 0\\ 0 & 0\end{pmatrix}V^*\mathbf{x} = \begin{pmatrix}\sigma & 0\\ 0 & 0\end{pmatrix}U^*\mathbf{y}. \tag{11.13}$$

One solution to this equation which is very easy to spot is

$$\mathbf{x} = V\begin{pmatrix}\sigma^{-1} & 0\\ 0 & 0\end{pmatrix}U^*\mathbf{y}. \tag{11.14}$$

Numerical Methods For Solving Linear Systems

12.0.1 Outcomes

A. Apply Gauss-Seidel iteration to approximate a solution to a linear system of equations.

B. Apply Jacobi iteration to approximate a solution to a linear system of equations.

12.1 Iterative Methods For Linear Systems

Consider the problem of solving the equation

$$A\mathbf{x} = \mathbf{b} \tag{12.1}$$

where A is an $n \times n$ matrix. In many applications, the matrix A is huge and composed mainly of zeros. For such matrices, the method of Gauss elimination (row operations) is not a good way to solve the system because the row operations can destroy the zeros and storing all those zeros takes a lot of room in a computer. These systems are called sparse. To solve them it is common to use an iterative technique. The idea is to obtain a sequence of approximate solutions which get close to the true solution after a sufficient number of iterations.

Definition 12.1.1 *Let $\{\mathbf{x}_k\}_{k=1}^{\infty}$ be a sequence of vectors in $\mathbb{F}^n$. Say*

$$\mathbf{x}_k = \left(x_1^k, \cdots, x_n^k\right).$$

Then this sequence is said to converge to the vector, $\mathbf{x} = (x_1, \cdots, x_n) \in \mathbb{F}^n$, written as

$$\lim_{k\to\infty} \mathbf{x}_k = \mathbf{x}$$

if for each $j = 1, 2, \cdots, n$,

$$\lim_{k\to\infty} x_j^k = x_j.$$

In words, the sequence converges if the entries of the vectors in the sequence converge to the corresponding entries of $\mathbf{x}$.

Example 12.1.2 *Consider $\mathbf{x}_k = \left(\sin(1/k), \frac{k^2}{1+k^2}, \ln\left(\frac{1+k^2}{k^2}\right)\right)$. Find $\lim_{k\to\infty} \mathbf{x}_k$.*

From the above definition, this limit is the vector, $(0, 1, 0)$ because

$$\lim_{k\to\infty} \sin(1/k) = 0, \ \lim_{k\to\infty} \frac{k^2}{1+k^2} = 1, \ \text{ and } \lim_{k\to\infty} \ln\left(\frac{1+k^2}{k^2}\right) = 0.$$

A more complete mathematical explanation is given in Linear Algebra.

12.1.1 The Jacobi Method

The first technique to be discussed here is the Jacobi method which is described in the following definition. In this technique, you have a sequence of vectors, $\{\mathbf{x}^k\}$ which converge to the solution to the linear system of equations and to get the i^{th} component of the $\mathbf{x}^{k+1}$, you use all the components of $\mathbf{x}^k$ except for the i^{th}. The precise description follows.

Definition 12.1.3 *The **Jacobi** iterative technique, also called the method of **simultaneous corrections,** is defined as follows. Let $\mathbf{x}^1$ be an initial vector, say the zero vector or some other vector. The method generates a succession of vectors, $\mathbf{x}^2, \mathbf{x}^3, \mathbf{x}^4, \cdots$ and hopefully this sequence of vectors will converge to the solution to 12.1. The vectors in this list are called **iterates** and they are obtained according to the following procedure. Letting $A = (a_{ij})$,*

$$a_{ii}x_i^{r+1} = -\sum_{j\neq i} a_{ij}x_j^r + b_i. \tag{12.2}$$

In terms of matrices, letting

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

The iterates are defined as

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & a_{nn} \end{pmatrix} \begin{pmatrix} x_1^{r+1} \\ x_2^{r+1} \\ \vdots \\ x_n^{r+1} \end{pmatrix}$$

$$= -\begin{pmatrix} 0 & a_{12} & \cdots & a_{1n} \\ a_{21} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n-1n} \\ a_{n1} & \cdots & a_{nn-1} & 0 \end{pmatrix} \begin{pmatrix} x_1^r \\ x_2^r \\ \vdots \\ x_n^r \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \tag{12.3}$$

The matrix on the left in 12.3 is obtained by retaining the main diagonal of A and setting every other entry equal to zero. The matrix on the right in 12.3 is obtained from A by setting every diagonal entry equal to zero and retaining all the other entries unchanged.

Example 12.1.4 *Use the Jacobi method to solve the system*

$$\begin{pmatrix} 3 & 1 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 2 & 5 & 1 \\ 0 & 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

In terms of the matrices, the Jacobi iteration is of the form

$$\begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} x_1^{r+1} \\ x_2^{r+1} \\ x_3^{r+1} \\ x_4^{r+1} \end{pmatrix} = -\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \\ 0 & 0 & 2 & 0 \end{pmatrix} \begin{pmatrix} x_1^r \\ x_2^r \\ x_3^r \\ x_4^r \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}.$$

Now iterate this starting with

$$\mathbf{x}^1 \equiv \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

$$\begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} x_1^2 \\ x_2^2 \\ x_3^2 \\ x_4^2 \end{pmatrix} = -\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \\ 0 & 0 & 2 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$
$$= \begin{pmatrix} 1.0 \\ 2.0 \\ 3.0 \\ 4.0 \end{pmatrix}$$

Solving this system yields

$$\mathbf{x}^2 = \begin{pmatrix} x_1^2 \\ x_2^2 \\ x_3^2 \\ x_4^2 \end{pmatrix} = \begin{pmatrix} .333\,333\,33 \\ .5 \\ .6 \\ 1.0 \end{pmatrix}$$

Then you use $\mathbf{x}^2$ to find $\mathbf{x}^3 = \begin{pmatrix} x_1^3 & x_2^3 & x_3^3 & x_4^3 \end{pmatrix}^T$

$$\begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} x_1^3 \\ x_2^3 \\ x_3^3 \\ x_4^3 \end{pmatrix} = -\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \\ 0 & 0 & 2 & 0 \end{pmatrix} \begin{pmatrix} .333\,333\,33 \\ .5 \\ .6 \\ 1.0 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$
$$= \begin{pmatrix} .5 \\ 1.066\,666\,7 \\ 1.0 \\ 2.8 \end{pmatrix}$$

The solution is

$$\mathbf{x}^3 = \begin{pmatrix} x_1^3 \\ x_2^3 \\ x_3^3 \\ x_4^3 \end{pmatrix} = \begin{pmatrix} .166\,666\,67 \\ .266\,666\,68 \\ .2 \\ .7 \end{pmatrix}$$

Now use this as the new data to find $\mathbf{x}^4 = \begin{pmatrix} x_1^4 & x_2^4 & x_3^4 & x_4^4 \end{pmatrix}^T$

$$\begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} x_1^4 \\ x_2^4 \\ x_3^4 \\ x_4^4 \end{pmatrix} = -\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \\ 0 & 0 & 2 & 0 \end{pmatrix} \begin{pmatrix} .166\,666\,67 \\ .266\,666\,68 \\ .2 \\ .7 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$
$$= \begin{pmatrix} .733\,333\,32 \\ 1.633\,333\,3 \\ 1.766\,666\,6 \\ 3.6 \end{pmatrix}.$$

Thus you find

$$\mathbf{x}^4 = \begin{pmatrix} .244\,444\,44 \\ .408\,333\,33 \\ .353\,333\,32 \\ .9 \end{pmatrix}$$

Then another iteration for $\mathbf{x}^5$ gives

$$\begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} x_1^5 \\ x_2^5 \\ x_3^5 \\ x_4^5 \end{pmatrix} = -\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \\ 0 & 0 & 2 & 0 \end{pmatrix} \begin{pmatrix} .244\,444\,44 \\ .408\,333\,33 \\ .353\,333\,32 \\ .9 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

$$= \begin{pmatrix} .591\,666\,67 \\ 1.402\,222\,2 \\ 1.283\,333\,3 \\ 3.293\,333\,4 \end{pmatrix}$$

and so

$$\mathbf{x}^5 = \begin{pmatrix} .197\,222\,22 \\ .350\,555\,55 \\ .256\,666\,66 \\ .823\,333\,35 \end{pmatrix}.$$

The solution to the system of equations obtained by row operations is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} .206 \\ .379 \\ .275 \\ .862 \end{pmatrix}$$

so already after only five iterations the iterates are pretty close to the true solution. How well does it work?

$$\begin{pmatrix} 3 & 1 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 2 & 5 & 1 \\ 0 & 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} .197\,222\,22 \\ .350\,555\,55 \\ .256\,666\,66 \\ .823\,333\,35 \end{pmatrix} = \begin{pmatrix} .942\,222\,21 \\ 1.856\,111\,1 \\ 2.807\,777\,8 \\ 3.806\,666\,7 \end{pmatrix}$$

$$\approx \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

A few more iterates will yield a better solution.

12.1.2 The Gauss Seidel Method

The Gauss Seidel method differs from the Jacobi method in using x_j^{k+1} for all $j < i$ in going from $\mathbf{x}^k$ to $\mathbf{x}^{k+1}$. This is why it is called the method of successive corrections. The precise description of this method is in the following definition.

Definition 12.1.5 *The **Gauss Seidel** method, also called the **method of successive corrections** is given as follows. For $A = (a_{ij})$, the iterates for the problem $A\mathbf{x} = \mathbf{b}$ are obtained according to the formula*

$$\sum_{j=1}^{i} a_{ij} x_j^{r+1} = -\sum_{j=i+1}^{n} a_{ij} x_j^r + b_i. \tag{12.4}$$

In terms of matrices, letting

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}$$

The iterates are defined as

$$\begin{aligned}\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{n1} & \cdots & a_{nn-1} & a_{nn} \end{pmatrix} \begin{pmatrix} x_1^{r+1} \\ x_2^{r+1} \\ \vdots \\ x_n^{r+1} \end{pmatrix} \\ = -\begin{pmatrix} 0 & a_{12} & \cdots & a_{1n} \\ 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a_{n-1n} \\ 0 & \cdots & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1^r \\ x_2^r \\ \vdots \\ x_n^r \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}\end{aligned} \tag{12.5}$$

In words, you set every entry in the original matrix which is strictly above the main diagonal equal to zero to obtain the matrix on the left. To get the matrix on the right, you set every entry of A which is on or below the main diagonal equal to zero. Using the iteration procedure of 12.4 directly, the Gauss Seidel method makes use of the very latest information which is available at that stage of the computation.

The following example is the same as the example used to illustrate the Jacobi method.

Example 12.1.6 *Use the Gauss Seidel method to solve the system*

$$\begin{pmatrix} 3 & 1 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 2 & 5 & 1 \\ 0 & 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

In terms of matrices, this procedure is

$$\begin{pmatrix} 3 & 0 & 0 & 0 \\ 1 & 4 & 0 & 0 \\ 0 & 2 & 5 & 0 \\ 0 & 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} x_1^{r+1} \\ x_2^{r+1} \\ x_3^{r+1} \\ x_4^{r+1} \end{pmatrix} = -\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1^r \\ x_2^r \\ x_3^r \\ x_4^r \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}.$$

As before, let $\mathbf{x}^1$ be the zero vector. Thus the first iteration is to solve

$$\begin{aligned}\begin{pmatrix} 3 & 0 & 0 & 0 \\ 1 & 4 & 0 & 0 \\ 0 & 2 & 5 & 0 \\ 0 & 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} x_1^2 \\ x_2^2 \\ x_3^2 \\ x_4^2 \end{pmatrix} &= -\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}\end{aligned}$$

Hence

$$\mathbf{x}^2 = \begin{pmatrix} .333\,333\,33 \\ .416\,666\,67 \\ .433\,333\,33 \\ .783\,333\,33 \end{pmatrix}$$

Thus $\mathbf{x}^3 = \begin{pmatrix} x_1^3 & x_2^3 & x_3^3 & x_4^3 \end{pmatrix}^T$ is given by

$$\begin{pmatrix} 3 & 0 & 0 & 0 \\ 1 & 4 & 0 & 0 \\ 0 & 2 & 5 & 0 \\ 0 & 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} x_1^3 \\ x_2^3 \\ x_3^3 \\ x_4^3 \end{pmatrix} = -\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} .333\,333\,33 \\ .416\,666\,67 \\ .433\,333\,33 \\ .783\,333\,33 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

$$= \begin{pmatrix} .583\,333\,33 \\ 1.566\,666\,7 \\ 2.216\,666\,7 \\ 4.0 \end{pmatrix}$$

And so

$$\mathbf{x}^3 = \begin{pmatrix} .194\,444\,44 \\ .343\,055\,56 \\ .306\,111\,11 \\ .846\,944\,44 \end{pmatrix}.$$

Another iteration for $\mathbf{x}^4$ involves solving

$$\begin{pmatrix} 3 & 0 & 0 & 0 \\ 1 & 4 & 0 & 0 \\ 0 & 2 & 5 & 0 \\ 0 & 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} x_1^4 \\ x_2^4 \\ x_3^4 \\ x_4^4 \end{pmatrix} = -\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} .194\,444\,44 \\ .343\,055\,56 \\ .306\,111\,11 \\ .846\,944\,44 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

$$= \begin{pmatrix} .656\,944\,44 \\ 1.693\,888\,9 \\ 2.153\,055\,6 \\ 4.0 \end{pmatrix}$$

and so

$$\mathbf{x}^4 = \begin{pmatrix} .218\,981\,48 \\ .368\,726\,86 \\ .283\,120\,38 \\ .858\,439\,81 \end{pmatrix}$$

Recall the answer is

$$\begin{pmatrix} .206 \\ .379 \\ .275 \\ .862 \end{pmatrix}$$

so the iterates are already pretty close to the answer. You could continue doing these iterates and it appears they converge to the solution. Now consider the following example.

Example 12.1.7 *Use the Gauss Seidel method to solve the system*

$$\begin{pmatrix} 1 & 4 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 2 & 5 & 1 \\ 0 & 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

The exact solution is given by doing row operations on the augmented matrix. When this is done the row echelon form is

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 6 \\ 0 & 1 & 0 & 0 & -\frac{5}{4} \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & \frac{1}{2} \end{pmatrix}$$

and so the solution is approximately

$$\begin{pmatrix} 6 \\ -\frac{5}{4} \\ 1 \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 6.0 \\ -1.25 \\ 1.0 \\ .5 \end{pmatrix}$$

The Gauss Seidel iterations are of the form

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 4 & 0 & 0 \\ 0 & 2 & 5 & 0 \\ 0 & 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} x_1^{r+1} \\ x_2^{r+1} \\ x_3^{r+1} \\ x_4^{r+1} \end{pmatrix} = - \begin{pmatrix} 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1^r \\ x_2^r \\ x_3^r \\ x_4^r \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

and so, multiplying by the inverse of the matrix on the left, the iteration reduces to the following in terms of matrix multiplication.

$$\mathbf{x}^{r+1} = - \begin{pmatrix} 0 & 4 & 0 & 0 \\ 0 & -1 & \frac{1}{4} & 0 \\ 0 & \frac{2}{5} & -\frac{1}{10} & \frac{1}{5} \\ 0 & -\frac{1}{5} & \frac{1}{20} & -\frac{1}{10} \end{pmatrix} \mathbf{x}^r + \begin{pmatrix} 1 \\ \frac{1}{4} \\ \frac{1}{2} \\ \frac{3}{4} \end{pmatrix}.$$

This time, we will pick an initial vector close to the answer. Let

$$\mathbf{x}^1 = \begin{pmatrix} 6 \\ -1 \\ 1 \\ \frac{1}{2} \end{pmatrix}$$

This is very close to the answer. Now lets see what the Gauss Seidel iteration does to it.

$$\mathbf{x}^2 = - \begin{pmatrix} 0 & 4 & 0 & 0 \\ 0 & -1 & \frac{1}{4} & 0 \\ 0 & \frac{2}{5} & -\frac{1}{10} & \frac{1}{5} \\ 0 & -\frac{1}{5} & \frac{1}{20} & -\frac{1}{10} \end{pmatrix} \begin{pmatrix} 6 \\ -1 \\ 1 \\ \frac{1}{2} \end{pmatrix} + \begin{pmatrix} 1 \\ \frac{1}{4} \\ \frac{1}{2} \\ \frac{3}{4} \end{pmatrix} = \begin{pmatrix} 5.0 \\ -1.0 \\ .9 \\ .55 \end{pmatrix}$$

You can't expect to be real close after only one iteration. Lets do another.

$$\mathbf{x}^3 = - \begin{pmatrix} 0 & 4 & 0 & 0 \\ 0 & -1 & \frac{1}{4} & 0 \\ 0 & \frac{2}{5} & -\frac{1}{10} & \frac{1}{5} \\ 0 & -\frac{1}{5} & \frac{1}{20} & -\frac{1}{10} \end{pmatrix} \begin{pmatrix} 5.0 \\ -1.0 \\ .9 \\ .55 \end{pmatrix} + \begin{pmatrix} 1 \\ \frac{1}{4} \\ \frac{1}{2} \\ \frac{3}{4} \end{pmatrix} = \begin{pmatrix} 5.0 \\ -.975 \\ .88 \\ .56 \end{pmatrix}$$

$$\mathbf{x}^4 = - \begin{pmatrix} 0 & 4 & 0 & 0 \\ 0 & -1 & \frac{1}{4} & 0 \\ 0 & \frac{2}{5} & -\frac{1}{10} & \frac{1}{5} \\ 0 & -\frac{1}{5} & \frac{1}{20} & -\frac{1}{10} \end{pmatrix} \begin{pmatrix} 5.0 \\ -.975 \\ .88 \\ .56 \end{pmatrix} + \begin{pmatrix} 1 \\ \frac{1}{4} \\ \frac{1}{2} \\ \frac{3}{4} \end{pmatrix} = \begin{pmatrix} 4.9 \\ -.945 \\ .866 \\ .567 \end{pmatrix}$$

The iterates seem to be getting farther from the actual solution. Why is the process which worked so well in the other examples not working here? A better question might be: Why does either process ever work at all?. A complete answer to this question is given in [9].

Both iterative procedures for solving

$$A\mathbf{x} = \mathbf{b} \tag{12.6}$$

are of the form

$$B\mathbf{x}^{r+1} = -C\mathbf{x}^r + \mathbf{b}$$

where $A = B + C$. In the Jacobi procedure, the matrix C was obtained by setting the diagonal of A equal to zero and leaving all other entries the same while the matrix, B was obtained by making every entry of A equal to zero other than the diagonal entries which are left unchanged. In the Gauss Seidel procedure, the matrix B was obtained from A by making every entry strictly above the main diagonal equal to zero and leaving the others unchanged and C was obtained from A by making every entry on or below the main diagonal equal to zero and leaving the others unchanged. Thus in the Jacobi procedure, B is a diagonal matrix while in the Gauss Seidel procedure, B is lower triangular. Using matrices to explicitly solve for the iterates, yields

$$\mathbf{x}^{r+1} = -B^{-1}C\mathbf{x}^r + B^{-1}\mathbf{b}. \tag{12.7}$$

This is what you would never have the computer do but this is what will allow the statement of a theorem which gives the condition for convergence of these and all other similar methods.

Theorem 12.1.8 *Let $A = B + C$ and suppose all eigenvalues of $B^{-1}C$ have absolute value less than 1 where $A = B + C$. Then the iterates in 12.7 converge to the unique solution of 12.6.*

A complete explanation of this important result is found in [9]. It depends on a theorem of Gelfand which is completely proved in this reference. Theorem 12.1.8 is very remarkable because it gives an algebraic condition for convergence which is essentially an analytical question.

Numerical Methods For Solving The Eigenvalue Problem

13.0.3 Outcomes

A. Apply the power method with scaling to approximate the dominant eigenvector corresponding to a dominant eigenvalue.

B. Use the shifted inverse power method to find the eigenvector and eigenvalue close to some number.

C. Approximate an eigenvalue of a symmetric matrix by computing the Rayleigh quotient and finding the associated error bound. Illustrate why the Rayleigh quotient approximates the dominant eigenvalue.

13.1 The Power Method For Eigenvalues

As indicated earlier, the eigenvalue eigenvector problem is extremely difficult. Consider for example what happens if you cannot find the eigenvalues exactly. Then you can't find an eigenvector because there isn't one due to the fact that $A - \lambda I$ is invertible whenever λ is not exactly equal to an eigenvalue. Therefore the straightforward way of solving this problem fails right away, even if you can approximate the eigenvalues. The power method allows you to approximate the largest eigenvalue and also the eigenvector which goes with it. By considering the inverse of the matrix, you can also find the smallest eigenvalue. The method works in the situation of a nondefective matrix, A which has an eigenvalue of algebraic multiplicity 1, λ_n which has the property that $|\lambda_k| < |\lambda_n|$ for all $k \neq n$. Note that for a real matrix this excludes the case that λ_n could be complex. Why? Such an eigenvalue is called a dominant eigenvalue.

Let $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ be a basis of eigenvectors for $\mathbb{F}^n$ such that $A\mathbf{x}_n = \lambda_n \mathbf{x}_n$. Now let $\mathbf{u}_1$ be some nonzero vector. Since $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ is a basis, there exists unique scalars, c_i such that

$$\mathbf{u}_1 = \sum_{k=1}^{n} c_k \mathbf{x}_k.$$

Assume you have not been so unlucky as to pick $\mathbf{u}_1$ in such a way that $c_n = 0$. Then let $A\mathbf{u}_k = \mathbf{u}_{k+1}$ so that

$$\mathbf{u}_m = A^m \mathbf{u}_1 = \sum_{k=1}^{n-1} c_k \lambda_k^m \mathbf{x}_k + \lambda_n^m c_n \mathbf{x}_n. \tag{13.1}$$

For large m the last term, $\lambda_n^m c_n \mathbf{x}_n$, determines quite well the direction of the vector on the right. This is because $|\lambda_n|$ is larger than $|\lambda_k|$ and so for a large, m, the sum, $\sum_{k=1}^{n-1} c_k \lambda_k^m \mathbf{x}_k$, on the right is fairly insignificant. Therefore, for large m, $\mathbf{u}_m$ is essentially a multiple of the eigenvector, $\mathbf{x}_n$, the one which goes with λ_n. The only problem is that there is no control of the size of the vectors $\mathbf{u}_m$. You can fix this by scaling. Let S_2 denote the entry of $A\mathbf{u}_1$ which is largest in absolute value. We call this a **scaling factor**. Then $\mathbf{u}_2$ will not be just $A\mathbf{u}_1$ but $A\mathbf{u}_1/S_2$. Next let S_3 denote the entry of $A\mathbf{u}_2$ which has largest absolute value and define $\mathbf{u}_3 \equiv A\mathbf{u}_2/S_3$. Continue this way. The scaling just described does not destroy the relative insignificance of the term involving a sum in 13.1. Indeed it amounts to nothing more than changing the units of length. Also note that from this scaling procedure, the absolute value of the largest element of $\mathbf{u}_k$ is always equal to 1. Therefore, for large m,

$$\mathbf{u}_m = \frac{\lambda_n^m c_n \mathbf{x}_n}{S_2 S_3 \cdots S_m} + (\text{relatively insignificant term}).$$

Therefore, the entry of $A\mathbf{u}_m$ which has the largest absolute value is essentially equal to the entry having largest absolute value of

$$A\left(\frac{\lambda_n^m c_n \mathbf{x}_n}{S_2 S_3 \cdots S_m}\right) = \frac{\lambda_n^{m+1} c_n \mathbf{x}_n}{S_2 S_3 \cdots S_m} \approx \lambda_n \mathbf{u}_m$$

and so for large m, it must be the case that $\lambda_n \approx S_{m+1}$. This suggests the following procedure.

Finding the largest eigenvalue with its eigenvector.

1. Start with a vector, $\mathbf{u}_1$ which you hope has a component in the direction of $\mathbf{x}_n$. The vector, $(1, \cdots, 1)^T$ is usually a pretty good choice.
2. If $\mathbf{u}_k$ is known,
$$\mathbf{u}_{k+1} = \frac{A\mathbf{u}_k}{S_{k+1}}$$
where S_{k+1} is the entry of $A\mathbf{u}_k$ which has largest absolute value.
3. When the scaling factors, S_k are not changing much, S_{k+1} will be close to the eigenvalue and $\mathbf{u}_{k+1}$ will be close to an eigenvector.
4. Check your answer to see if it worked well.

Example 13.1.1 *Find the largest eigenvalue of* $A = \begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix}$.

The power method will now be applied to find the largest eigenvalue for the above matrix. Letting $\mathbf{u}_1 = (1, \cdots, 1)^T$, we will consider $A\mathbf{u}_1$ and scale it.

$$\begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ -4 \\ 6 \end{pmatrix}.$$

Scaling this vector by dividing by the largest entry gives

$$\frac{1}{6}\begin{pmatrix} 2 \\ -4 \\ 6 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ -\frac{2}{3} \\ 1 \end{pmatrix} = \mathbf{u}_2$$

Now lets do it again.

$$\begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix} \begin{pmatrix} \frac{1}{3} \\ -\frac{2}{3} \\ 1 \end{pmatrix} = \begin{pmatrix} 22 \\ -8 \\ -6 \end{pmatrix}$$

Then

$$\mathbf{u}_3 = \frac{1}{22} \begin{pmatrix} 22 \\ -8 \\ -6 \end{pmatrix} = \begin{pmatrix} 1 \\ -\frac{4}{11} \\ -\frac{3}{11} \end{pmatrix} = \begin{pmatrix} 1.0 \\ -.363\,636\,36 \\ -.272\,727\,27 \end{pmatrix}.$$

Continue doing this

$$\begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix} \begin{pmatrix} 1.0 \\ -.363\,636\,36 \\ -.272\,727\,27 \end{pmatrix} = \begin{pmatrix} 7.090\,909\,1 \\ -4.363\,636\,4 \\ 1.636\,363\,7 \end{pmatrix}$$

Then

$$\mathbf{u}_4 = \begin{pmatrix} 1.0 \\ -.615\,38 \\ .230\,77 \end{pmatrix}$$

So far the scaling factors are changing fairly noticeably so continue.

$$\begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix} \begin{pmatrix} 1.0 \\ -.615\,38 \\ .230\,77 \end{pmatrix} = \begin{pmatrix} 16.154 \\ -7.384\,6 \\ -1.384\,6 \end{pmatrix}$$

$$\mathbf{u}_5 = \begin{pmatrix} 1.0 \\ -.457\,14 \\ -8.571\,3 \times 10^{-2} \end{pmatrix}$$

$$\begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix} \begin{pmatrix} 1.0 \\ -.457\,14 \\ -8.571\,3 \times 10^{-2} \end{pmatrix} = \begin{pmatrix} 10.457 \\ -5.485\,7 \\ .514\,3 \end{pmatrix}$$

$$\mathbf{u}_6 = \begin{pmatrix} 1.0 \\ -.524\,6 \\ 4.918\,2 \times 10^{-2} \end{pmatrix}$$

$$\begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix} \begin{pmatrix} 1.0 \\ -.524\,6 \\ 4.918\,2 \times 10^{-2} \end{pmatrix} = \begin{pmatrix} 12.885 \\ -6.295\,1 \\ -.295\,15 \end{pmatrix}$$

$$\mathbf{u}_7 = \begin{pmatrix} 1.0 \\ -.488\,56 \\ -2.290\,6 \times 10^{-2} \end{pmatrix}$$

$$\begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix} \begin{pmatrix} 1.0 \\ -.488\,56 \\ -2.290\,6 \times 10^{-2} \end{pmatrix} = \begin{pmatrix} 11.588 \\ -5.862\,6 \\ .137\,36 \end{pmatrix}$$

$$\mathbf{u}_8 = \begin{pmatrix} 1.0 \\ -.505\,92 \\ 1.185\,4 \times 10^{-2} \end{pmatrix}$$

$$\begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix} \begin{pmatrix} 1.0 \\ -.505\,92 \\ 1.185\,4 \times 10^{-2} \end{pmatrix} = \begin{pmatrix} 12.213 \\ -6.071\,1 \\ -7.108\,2 \times 10^{-2} \end{pmatrix}$$

$$\mathbf{u}_9 = \begin{pmatrix} 1.0 \\ -.497\,1 \\ -5.820\,2 \times 10^{-3} \end{pmatrix}$$

$$\begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix} \begin{pmatrix} 1.0 \\ -.497\,1 \\ -5.820\,2 \times 10^{-3} \end{pmatrix} = \begin{pmatrix} 11.895 \\ -5.965\,1 \\ 3.486\,1 \times 10^{-2} \end{pmatrix}$$

$$\mathbf{u}_{10} = \begin{pmatrix} 1.0 \\ -.501\,48 \\ 2.930\,7 \times 10^{-3} \end{pmatrix}$$

$$\begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix} \begin{pmatrix} 1.0 \\ -.501\,48 \\ 2.930\,7 \times 10^{-3} \end{pmatrix} = \begin{pmatrix} 12.053 \\ -6.017\,6 \\ -1.767\,2 \times 10^{-2} \end{pmatrix}$$

$$\mathbf{u}_{11} = \begin{pmatrix} 1.0 \\ -.499\,26 \\ -1.466\,2 \times 10^{-3} \end{pmatrix}$$

At this point, you could stop because the scaling factors are not changing by much. They went from 11.895 to 12.053. It looks like the eigenvalue is something like 12 which is in fact the case. The eigenvector is approximately $\mathbf{u}_{11}$. The true eigenvector for $\lambda = 12$ is

$$\begin{pmatrix} 1 \\ -.5 \\ 0 \end{pmatrix}$$

and so you see this is pretty close. If you didn't know this, observe

$$\begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix} \begin{pmatrix} 1.0 \\ -.499\,26 \\ -1.466\,2 \times 10^{-3} \end{pmatrix} = \begin{pmatrix} 11.974 \\ -5.991\,2 \\ 8.838\,6 \times 10^{-3} \end{pmatrix}$$

and

$$12.053 \begin{pmatrix} 1.0 \\ -.499\,26 \\ -1.466\,2 \times 10^{-3} \end{pmatrix} = \begin{pmatrix} 12.053 \\ -6.017\,6 \\ -1.767\,2 \times 10^{-2} \end{pmatrix}.$$

13.2 The Shifted Inverse Power Method

This method can find various eigenvalues and eigenvectors. It is a significant generalization of the above simple procedure and yields very good results. The situation is this: You have a number, α which is close to λ, some eigenvalue of an $n \times n$ matrix, A. You don't know λ but you know that α is closer to λ than to any other eigenvalue. Your problem is to find both λ and an eigenvector which goes with λ. Another way to look at this is to start with α and seek the eigenvalue, λ, which is closest to α along with an eigenvector associated with λ. If α is an eigenvalue of A, then you have what you want. Therefore, we will always assume α is not an eigenvalue of A and so $(A - \alpha I)^{-1}$ exists. The method is based on the following lemma. When using this method it is nice to choose α fairly close to an eigenvalue.

Otherwise, the method will converge slowly. In order to get some idea where to start, you could use Gerschgorin's theorem but this theorem will only give a rough idea where to look. There isn't a really good way to know how to choose α for general cases. As we mentioned earlier, the eigenvalue problem is very difficult to solve in general.

Lemma 13.2.1 *Let $\{\lambda_k\}_{k=1}^n$ be the eigenvalues of A. If $\mathbf{x}_k$ is an eigenvector of A for the eigenvalue λ_k, then $\mathbf{x}_k$ is an eigenvector for $(A-\alpha I)^{-1}$ corresponding to the eigenvalue $\frac{1}{\lambda_k-\alpha}$.*

Proof: Let λ_k and $\mathbf{x}_k$ be as described in the statement of the lemma. Then

$$(A-\alpha I)\,\mathbf{x}_k = (\lambda_k-\alpha)\,\mathbf{x}_k$$

and so

$$\frac{1}{\lambda_k-\alpha}\mathbf{x}_k = (A-\alpha I)^{-1}\,\mathbf{x}_k.$$

This proves the lemma.

In explaining why the method works, we will assume A is nondefective. **This is not necessary!** One can use Gelfand's theorem on the spectral radius which is presented in [9] and invariance of $(A-\alpha I)^{-1}$ on generalized eigenspaces to prove more general results. It suffices to assume that the eigenspace for λ_k has dimension equal to the multiplicity of the eigenvalue λ_k but even this is not necessary to obtain convergence of the method. This method is better than might be supposed from the following explanation.

Pick $\mathbf{u}_1$, an initial vector and let $A\mathbf{x}_k = \lambda_k\mathbf{x}_k$, where $\{\mathbf{x}_1,\cdots,\mathbf{x}_n\}$ is a basis of eigenvectors which exists from the assumption that A is nondefective. Assume α is closer to λ_n than to any other eigenvalue. Since A is nondefective, there exist constants, a_k such that

$$\mathbf{u}_1 = \sum_{k=1}^n a_k\mathbf{x}_k.$$

Possibly λ_n is a repeated eigenvalue. Then combining the terms in the sum which involve eigenvectors for λ_n, a simpler description of $\mathbf{u}_1$ is

$$\mathbf{u}_1 = \sum_{j=1}^m a_j\mathbf{x}_j + \mathbf{y}$$

where $\mathbf{y}$ is an eigenvector for λ_n which is assumed not equal to $\mathbf{0}$. (If you are unlucky in your choice for $\mathbf{u}_1$, this might not happen and things won't work.) Now the iteration procedure is defined as

$$\mathbf{u}_{k+1} \equiv \frac{(A-\alpha I)^{-1}\,\mathbf{u}_k}{S_k}$$

where S_k is the element of $(A-\alpha I)^{-1}\,\mathbf{u}_k$ which has largest absolute value. From Lemma 13.2.1,

$$\begin{aligned}\mathbf{u}_{k+1} &= \frac{\sum_{j=1}^m a_j\left(\frac{1}{\lambda_j-\alpha}\right)^k\mathbf{x}_j + \left(\frac{1}{\lambda_n-\alpha}\right)^k\mathbf{y}}{S_2\cdots S_k}\\ &= \frac{\left(\frac{1}{\lambda_n-\alpha}\right)^k}{S_2\cdots S_k}\left(\sum_{j=1}^m a_j\left(\frac{\lambda_n-\alpha}{\lambda_j-\alpha}\right)^k\mathbf{x}_j + \mathbf{y}\right).\end{aligned}$$

Now it is being assumed that λ_n is the eigenvalue which is closest to α and so for large k, the term,

$$\sum_{j=1}^{m} a_j \left(\frac{\lambda_n - \alpha}{\lambda_j - \alpha} \right)^k \mathbf{x}_j \equiv \mathbf{E}_k$$

is very small while for every $k \geq 1$, $\mathbf{u}_k$ is a moderate sized vector because every entry has absolute value less than or equal to 1. Thus

$$\mathbf{u}_{k+1} = \frac{\left(\frac{1}{\lambda_n - \alpha} \right)^k}{S_2 \cdots S_k} (\mathbf{E}_k + \mathbf{y}) \equiv C_k (\mathbf{E}_k + \mathbf{y})$$

where $\mathbf{E}_k \to \mathbf{0}$, $\mathbf{y}$ is some eigenvector for λ_n, and C_k is of moderate size, remaining bounded as $k \to \infty$. Therefore, for large k,

$$\mathbf{u}_{k+1} - C_k \mathbf{y} = C_k \mathbf{E}_k \approx \mathbf{0}$$

and multiplying by $(A - \alpha I)^{-1}$ yields

$$\begin{aligned} (A - \alpha I)^{-1} \mathbf{u}_{k+1} - (A - \alpha I)^{-1} C_k \mathbf{y} &= (A - \alpha I)^{-1} \mathbf{u}_{k+1} - C_k \left(\frac{1}{\lambda_n - \alpha} \right) \mathbf{y} \\ &\approx (A - \alpha I)^{-1} \mathbf{u}_{k+1} - \left(\frac{1}{\lambda_n - \alpha} \right) \mathbf{u}_{k+1} \approx \mathbf{0}. \end{aligned}$$

Therefore, for large k, $\mathbf{u}_k$ is approximately equal to an eigenvector of $(A - \alpha I)^{-1}$. Therefore,

$$(A - \alpha I)^{-1} \mathbf{u}_k \approx \frac{1}{\lambda_n - \alpha} \mathbf{u}_k$$

and so you could take the dot product of both sides with $\mathbf{u}_k$ and approximate λ_n by solving the following for λ_n.

$$\frac{(A - \alpha I)^{-1} \mathbf{u}_k \cdot \mathbf{u}_k}{|\mathbf{u}_k|^2} = \frac{1}{\lambda_n - \alpha}$$

How else can you find the eigenvalue from this? Suppose $\mathbf{u}_k = (w_1, \cdots, w_n)^T$ and from the construction $|w_i| \leq 1$ and $w_k = 1$ for some k. Then

$$S_k \mathbf{u}_{k+1} = (A - \alpha I)^{-1} \mathbf{u}_k \approx (A - \alpha I)^{-1} (C_{k-1} \mathbf{y}) = \frac{1}{\lambda_n - \alpha} (C_{k-1} \mathbf{y}) \approx \frac{1}{\lambda_n - \alpha} \mathbf{u}_k.$$

Hence the entry of $(A - \alpha I)^{-1} \mathbf{u}_k$ which has largest absolute value is approximately $\frac{1}{\lambda_n - \alpha}$ and so it is likely that you can estimate λ_n using the formula

$$S_k = \frac{1}{\lambda_n - \alpha}.$$

Of course this would fail if $(A - \alpha I)^{-1} \mathbf{u}_k$ had more than one entry having equal absolute value.

Here is how you use the shifted inverse power method to find the eigenvalue and eigenvector closest to α.

1. Find $(A - \alpha I)^{-1}$.

2. Pick $\mathbf{u}_1$. It is important that $\mathbf{u}_1 = \sum_{j=1}^{m} a_j \mathbf{x}_j + \mathbf{y}$ where $\mathbf{y}$ is an eigenvector which goes with the eigenvalue closest to α and the sum is in an "invariant subspace corresponding to the other eigenvalues". Of course you have no way of knowing whether this is so but it typically is so. If things don't work out, just start with a different $\mathbf{u}_1$. You were unlucky in your choice.

3. If $\mathbf{u}_k$ has been obtained,
$$\mathbf{u}_{k+1} = \frac{(A - \alpha I)^{-1} \mathbf{u}_k}{S_k}$$
where S_k is the element of $\mathbf{u}_k$ which has largest absolute value.

4. When the scaling factors, S_k are not changing much and the $\mathbf{u}_k$ are not changing much, find the approximation to the eigenvalue by solving
$$S_k = \frac{1}{\lambda - \alpha}$$
for λ. The eigenvector is approximated by $\mathbf{u}_{k+1}$.

5. Check your work by multiplying by the original matrix to see how well what you have found works.

Example 13.2.2 *Find the eigenvalue of* $A = \begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix}$ *which is closest to* -7. *Also find an eigenvector which goes with this eigenvalue.*

In this case the eigenvalues are $-6, 0$, and 12 so the correct answer is -6 for the eigenvalue. Then from the above procedure, we will start with an initial vector,
$$\mathbf{u}_1 \equiv \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$
Then we must solve the following equation.
$$\left(\left(\begin{pmatrix} 5 & -14 & 11 \\ -4 & 4 & -4 \\ 3 & 6 & -3 \end{pmatrix} + 7 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right.$$
Simplifying the matrix on the left, we must solve
$$\begin{pmatrix} 12 & -14 & 11 \\ -4 & 11 & -4 \\ 3 & 6 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$
and then divide by the entry which has largest absolute value to obtain
$$\mathbf{u}_2 = \begin{pmatrix} 1.0 \\ .184 \\ -.76 \end{pmatrix}$$
Now solve
$$\begin{pmatrix} 12 & -14 & 11 \\ -4 & 11 & -4 \\ 3 & 6 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1.0 \\ .184 \\ -.76 \end{pmatrix}$$

and divide by the largest entry, 1.051 5 to get

$$\mathbf{u}_3 = \begin{pmatrix} 1.0 \\ .0\,266 \\ -.\,970\,61 \end{pmatrix}$$

Solve

$$\begin{pmatrix} 12 & -14 & 11 \\ -4 & 11 & -4 \\ 3 & 6 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1.0 \\ .0\,266 \\ -.\,970\,61 \end{pmatrix}$$

and divide by the largest entry, 1.01 to get

$$\mathbf{u}_4 = \begin{pmatrix} 1.0 \\ 3.\,845\,4 \times 10^{-3} \\ -.\,996\,04 \end{pmatrix}.$$

These scaling factors are pretty close after these few iterations. Therefore, the predicted eigenvalue is obtained by solving the following for λ.

$$\frac{1}{\lambda + 7} = 1.01$$

which gives $\lambda = -6.01$. You see this is pretty close. In this case the eigenvalue closest to -7 was -6.

Example 13.2.3 *Consider the symmetric matrix,* $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 2 \end{pmatrix}$. *Find the middle eigenvalue and an eigenvector which goes with it.*

Since A is symmetric, it follows it has three real eigenvalues which are solutions to

$$\begin{aligned} p(\lambda) &= \det\left(\lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 2 \end{pmatrix}\right) \\ &= \lambda^3 - 4\lambda^2 - 24\lambda - 17 = 0 \end{aligned}$$

If you use your graphing calculator to graph this polynomial, you find there is an eigenvalue somewhere between $-.9$ and $-.8$ and that this is the middle eigenvalue. Of course you could zoom in and find it very accurately without much trouble but what about the eigenvector which goes with it? If you try to solve

$$\left((-.8)\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 2 \end{pmatrix}\right) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

there will be only the zero solution because the matrix on the left will be invertible and the same will be true if you replace $-.8$ with a better approximation like $-.86$ or $-.855$. This is because all these are only approximations to the eigenvalue and so the matrix in the above is nonsingular for all of these. Therefore, you will only get the zero solution and

Eigenvectors are never equal to zero!

However, there exists such an eigenvector and you can find it using the shifted inverse power method. Pick $\alpha = -.855$. Then you solve

$$\left(\left(\begin{array}{ccc} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 2 \end{array}\right) + .855 \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right)\right) \left(\begin{array}{c} x \\ y \\ z \end{array}\right) = \left(\begin{array}{c} 1 \\ 1 \\ 1 \end{array}\right)$$

or in other words,

$$\left(\begin{array}{ccc} 1.855 & 2.0 & 3.0 \\ 2.0 & 1.855 & 4.0 \\ 3.0 & 4.0 & 2.855 \end{array}\right) \left(\begin{array}{c} x \\ y \\ z \end{array}\right) = \left(\begin{array}{c} 1 \\ 1 \\ 1 \end{array}\right).$$

Divide by the largest entry, -67.944, to obtain

$$\mathbf{u}_2 = \left(\begin{array}{c} 1.0 \\ -.589\,21 \\ -.230\,44 \end{array}\right).$$

Now solve

$$\left(\begin{array}{ccc} 1.855 & 2.0 & 3.0 \\ 2.0 & 1.855 & 4.0 \\ 3.0 & 4.0 & 2.855 \end{array}\right) \left(\begin{array}{c} x \\ y \\ z \end{array}\right) = \left(\begin{array}{c} 1.0 \\ -.589\,21 \\ -.230\,44 \end{array}\right).$$

The solution is : $\left(\begin{array}{c} -514.01 \\ 302.12 \\ 116.75 \end{array}\right)$ and divide by the largest entry, -514.01, to obtain

$$\mathbf{u}_3 = \left(\begin{array}{c} 1.0 \\ -.587\,77 \\ -.227\,14 \end{array}\right). \tag{13.2}$$

Clearly the $\mathbf{u}_k$ are not changing much. This suggests an approximate eigenvector for this eigenvalue which is close to $-.855$ is the above $\mathbf{u}_3$. And an eigenvalue is obtained by solving

$$\frac{1}{\lambda + .855} = -514.01$$

$\lambda = -.856\,9$. Lets check this.

$$\left(\begin{array}{ccc} 1 & 2 & 3 \\ 2 & 1 & 4 \\ 3 & 4 & 2 \end{array}\right) \left(\begin{array}{c} 1.0 \\ -.587\,77 \\ -.227\,14 \end{array}\right) = \left(\begin{array}{c} -.856\,96 \\ .503\,67 \\ .194\,64 \end{array}\right).$$

$$-.856\,9 \left(\begin{array}{c} 1.0 \\ -.587\,77 \\ -.227\,14 \end{array}\right) = \left(\begin{array}{c} -.856\,9 \\ .503\,7 \\ .194\,6 \end{array}\right)$$

Thus the vector of 13.2 is very close to the desired eigenvector, just as $-.856\,9$ is very close to the desired eigenvalue. For practical purposes, we have found both the eigenvector and the eigenvalue.

Example 13.2.4 *Find the eigenvalues and eigenvectors of the matrix,* $A = \left(\begin{array}{ccc} 2 & 1 & 3 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{array}\right)$.

This is only a 3×3 matrix and so it is not hard to estimate the eigenvalues. Just get the characteristic equation, graph it using a calculator and zoom in to find the eigenvalues. If you do this, you find there is an eigenvalue near -1.2, one near $-.4$, and one near 5.5. (The characteristic equation is $2+8\lambda+4\lambda^2-\lambda^3=0$.) Of course we have no idea what the eigenvectors are.

Lets first try to find the eigenvector and a better approximation for the eigenvalue near -1.2. In this case, let $\alpha=-1.2$. Then

$$(A-\alpha I)^{-1}=\begin{pmatrix} -25.357\,143 & -33.928\,571 & 50.0 \\ 12.5 & 17.5 & -25.0 \\ 23.214\,286 & 30.357\,143 & -45.0 \end{pmatrix}.$$

Then for the first iteration, letting $\mathbf{u}_1=(1,1,1)^T$,

$$\begin{pmatrix} -25.357\,143 & -33.928\,571 & 50.0 \\ 12.5 & 17.5 & -25.0 \\ 23.214\,286 & 30.357\,143 & -45.0 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}=\begin{pmatrix} -9.285\,714 \\ 5.0 \\ 8.571\,429 \end{pmatrix}$$

To get $\mathbf{u}_2$, we must divide by $-9.285\,714$. Thus

$$\mathbf{u}_2=\begin{pmatrix} 1.0 \\ -.538\,461\,56 \\ -.923\,077 \end{pmatrix}.$$

Do another iteration.

$$\begin{pmatrix} -25.357\,143 & -33.928\,571 & 50.0 \\ 12.5 & 17.5 & -25.0 \\ 23.214\,286 & 30.357\,143 & -45.0 \end{pmatrix}\begin{pmatrix} 1.0 \\ -.538\,461\,56 \\ -.923\,077 \end{pmatrix}=\begin{pmatrix} -53.241\,762 \\ 26.153\,848 \\ 48.406\,596 \end{pmatrix}$$

Then to get $\mathbf{u}_3$ you divide by $-53.241\,762$. Thus

$$\mathbf{u}_3=\begin{pmatrix} 1.0 \\ -.491\,228\,07 \\ -.909\,184\,71 \end{pmatrix}.$$

Now iterate again because the scaling factors are still changing quite a bit.

$$\begin{pmatrix} -25.357\,143 & -33.928\,571 & 50.0 \\ 12.5 & 17.5 & -25.0 \\ 23.214\,286 & 30.357\,143 & -45.0 \end{pmatrix}\begin{pmatrix} 1.0 \\ -.491\,228\,07 \\ -.909\,184\,71 \end{pmatrix}=\begin{pmatrix} -54.149\,712 \\ 26.633\,127 \\ 49.215\,317 \end{pmatrix}.$$

This time the scaling factor didn't change too much. It is $-54.149\,712$. Thus

$$\mathbf{u}_4=\begin{pmatrix} 1.0 \\ -.491\,842\,45 \\ -.908\,874\,95 \end{pmatrix}.$$

Lets do one more iteration.

$$\begin{pmatrix} -25.357\,143 & -33.928\,571 & 50.0 \\ 12.5 & 17.5 & -25.0 \\ 23.214\,286 & 30.357\,143 & -45.0 \end{pmatrix}\begin{pmatrix} 1.0 \\ -.491\,842\,45 \\ -.908\,874\,95 \end{pmatrix}=\begin{pmatrix} -54.113\,379 \\ 26.614\,631 \\ 49.182\,727 \end{pmatrix}.$$

You see at this point the scaling factors have definitely settled down and so it seems our eigenvalue would be obtained by solving

$$\frac{1}{\lambda-(-1.2)}=-54.113\,379$$

and this yields $\lambda = -1.218\,479\,7$ as an approximation to the eigenvalue and the eigenvector would be obtained by dividing by $-54.113\,379$ which gives

$$\mathbf{u}_5 = \begin{pmatrix} 1.000\,000 \\ -.491\,830\,9 \\ -.908\,883\,0 \end{pmatrix}.$$

How well does it work?

$$\begin{pmatrix} 2 & 1 & 3 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1.000\,000 \\ -.491\,830\,9 \\ -.908\,883\,0 \end{pmatrix} = \begin{pmatrix} -1.218\,5 \\ .599\,29 \\ 1.107\,5 \end{pmatrix}$$

while

$$-1.218\,479\,7 \begin{pmatrix} 1.000\,000 \\ -.491\,830\,9 \\ -.908\,883\,0 \end{pmatrix} = \begin{pmatrix} -1.218\,5 \\ .599\,29 \\ 1.107\,5 \end{pmatrix}.$$

For practical purposes, this has found the eigenvalue near -1.2 as well as an eigenvector associated with it.

Next we shall find the eigenvector and a more precise value for the eigenvalue near $-.4$. In this case,

$$(A - \alpha I)^{-1} = \begin{pmatrix} 8.064\,516\,1 \times 10^{-2} & -9.274\,193\,5 & 6.451\,612\,9 \\ -.403\,225\,81 & 11.370\,968 & -7.258\,064\,5 \\ .403\,225\,81 & 3.629\,032\,3 & -2.741\,935\,5 \end{pmatrix}.$$

As before, we have no idea what the eigenvector is so we will again try $(1,1,1)^T$. Then to find $\mathbf{u}_2$,

$$\begin{pmatrix} 8.064\,516\,1 \times 10^{-2} & -9.274\,193\,5 & 6.451\,612\,9 \\ -.403\,225\,81 & 11.370\,968 & -7.258\,064\,5 \\ .403\,225\,81 & 3.629\,032\,3 & -2.741\,935\,5 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -2.741\,935\,4 \\ 3.709\,677\,7 \\ 1.290\,322\,6 \end{pmatrix}$$

The scaling factor is $3.709\,677\,7$. Thus

$$\mathbf{u}_2 = \begin{pmatrix} -.739\,130\,36 \\ 1.0 \\ .347\,826\,07 \end{pmatrix}.$$

Now lets do another iteration.

$$\begin{pmatrix} 8.064\,516\,1 \times 10^{-2} & -9.274\,193\,5 & 6.451\,612\,9 \\ -.403\,225\,81 & 11.370\,968 & -7.258\,064\,5 \\ .403\,225\,81 & 3.629\,032\,3 & -2.741\,935\,5 \end{pmatrix} \begin{pmatrix} -.739\,130\,36 \\ 1.0 \\ .347\,826\,07 \end{pmatrix}$$

$$= \begin{pmatrix} -7.089\,761\,6 \\ 9.144\,460\,4 \\ 2.377\,279\,2 \end{pmatrix}.$$

The scaling factor is $9.144\,460\,4$. Thus

$$\mathbf{u}_3 = \begin{pmatrix} -.775\,306\,72 \\ 1.0 \\ .259\,969\,33 \end{pmatrix}.$$

Lets do another iteration. The scaling factors are still changing quite a bit.

$$\begin{pmatrix} 8.064\,516\,1\times10^{-2} & -9.274\,193\,5 & 6.451\,612\,9 \\ -.403\,225\,81 & 11.370\,968 & -7.258\,064\,5 \\ .403\,225\,81 & 3.629\,032\,3 & -2.741\,935\,5 \end{pmatrix} \begin{pmatrix} -.775\,306\,72 \\ 1.0 \\ .259\,969\,33 \end{pmatrix}$$
$$= \begin{pmatrix} -7.659\,496\,8 \\ 9.796\,717\,5 \\ 2.603\,589\,5 \end{pmatrix}.$$

The scaling factor is now $9.796\,717\,5$. Therefore,

$$\mathbf{u}_4 = \begin{pmatrix} -.781\,843\,18 \\ 1.0 \\ .265\,761\,41 \end{pmatrix}.$$

Lets do another iteration.

$$\begin{pmatrix} 8.064\,516\,1\times10^{-2} & -9.274\,193\,5 & 6.451\,612\,9 \\ -.403\,225\,81 & 11.370\,968 & -7.258\,064\,5 \\ .403\,225\,81 & 3.629\,032\,3 & -2.741\,935\,5 \end{pmatrix} \begin{pmatrix} -.781\,843\,18 \\ 1.0 \\ .265\,761\,41 \end{pmatrix}$$
$$= \begin{pmatrix} -7.622\,655\,6 \\ 9.757\,313\,9 \\ 2.585\,072\,3 \end{pmatrix}.$$

Now the scaling factor is $9.757\,313\,9$ and so

$$\mathbf{u}_5 = \begin{pmatrix} -.781\,224\,8 \\ 1.0 \\ .264\,936\,88 \end{pmatrix}.$$

We notice the scaling factors are not changing by much so the approximate eigenvalue is

$$\frac{1}{\lambda + .4} = 9.757\,313\,9$$

which shows $\lambda = -.297\,512\,78$ is an approximation to the eigenvalue near .4. How well does it work?

$$\begin{pmatrix} 2 & 1 & 3 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} -.781\,224\,8 \\ 1.0 \\ .264\,936\,88 \end{pmatrix} = \begin{pmatrix} .232\,361\,04 \\ -.297\,512\,72 \\ -.0\,787\,375\,2 \end{pmatrix}.$$

$$-.297\,512\,78 \begin{pmatrix} -.781\,224\,8 \\ 1.0 \\ .264\,936\,88 \end{pmatrix} = \begin{pmatrix} .232\,424\,36 \\ -.297\,512\,78 \\ -7.882\,210\,8\times10^{-2} \end{pmatrix}.$$

It works pretty well. For practical purposes, the eigenvalue and eigenvector have now been found. If you want better accuracy, you could just continue iterating.

Next we will find the eigenvalue and eigenvector for the eigenvalue near 5.5. In this case,

$$(A-\alpha I)^{-1} = \begin{pmatrix} 29.2 & 16.8 & 23.2 \\ 19.2 & 10.8 & 15.2 \\ 28.0 & 16.0 & 22.0 \end{pmatrix}.$$

As before, we have no idea what the eigenvector is but we don't want to give the impression that you always need to start with the vector $(1,1,1)^T$. Therefore, we shall let $\mathbf{u}_1 =$

$(1,2,3)^T$. What follows is the iteration without all the comments between steps.

$$\begin{pmatrix} 29.2 & 16.8 & 23.2 \\ 19.2 & 10.8 & 15.2 \\ 28.0 & 16.0 & 22.0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1.324 \times 10^2 \\ 86.4 \\ 1.26 \times 10^2 \end{pmatrix}.$$

$S_1 = 86.4$.

$$\mathbf{u}_2 = \begin{pmatrix} 1.532\,407\,4 \\ 1.0 \\ 1.458\,333\,3 \end{pmatrix}.$$

$$\begin{pmatrix} 29.2 & 16.8 & 23.2 \\ 19.2 & 10.8 & 15.2 \\ 28.0 & 16.0 & 22.0 \end{pmatrix} \begin{pmatrix} 1.532\,407\,4 \\ 1.0 \\ 1.458\,333\,3 \end{pmatrix} = \begin{pmatrix} 95.379\,629 \\ 62.388\,888 \\ 90.990\,74 \end{pmatrix}$$

$S_2 = 95.379\,629$.

$$\mathbf{u}_3 = \begin{pmatrix} 1.0 \\ .654\,111\,25 \\ .953\,985\,05 \end{pmatrix}$$

$$\begin{pmatrix} 29.2 & 16.8 & 23.2 \\ 19.2 & 10.8 & 15.2 \\ 28.0 & 16.0 & 22.0 \end{pmatrix} \begin{pmatrix} 1.0 \\ .654\,111\,25 \\ .953\,985\,05 \end{pmatrix} = \begin{pmatrix} 62.321\,522 \\ 40.764\,974 \\ 59.453\,451 \end{pmatrix}$$

$S_3 = 62.321\,522$.

$$\mathbf{u}_4 = \begin{pmatrix} 1.0 \\ .654\,107\,48 \\ .953\,979\,45 \end{pmatrix}$$

$$\begin{pmatrix} 29.2 & 16.8 & 23.2 \\ 19.2 & 10.8 & 15.2 \\ 28.0 & 16.0 & 22.0 \end{pmatrix} \begin{pmatrix} 1.0 \\ .654\,107\,48 \\ .953\,979\,45 \end{pmatrix} = \begin{pmatrix} 62.321\,329 \\ 40.764\,848 \\ 59.453\,268 \end{pmatrix}$$

$S_4 = 62.321\,329$. Looks like it is time to stop because this scaling factor is not changing much from S_3.

$$\mathbf{u}_5 = \begin{pmatrix} 1.0 \\ .654\,107\,49 \\ .953\,979\,46 \end{pmatrix}.$$

Then the approximation of the eigenvalue is gotten by solving

$$62.321\,329 = \frac{1}{\lambda - 5.5}$$

which gives $\lambda = 5.516\,045\,9$. Lets see how well it works.

$$\begin{pmatrix} 2 & 1 & 3 \\ 2 & 1 & 1 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1.0 \\ .654\,107\,49 \\ .953\,979\,46 \end{pmatrix} = \begin{pmatrix} 5.516\,045\,9 \\ 3.608\,087 \\ 5.262\,194\,4 \end{pmatrix}$$

$$5.516\,045\,9 \begin{pmatrix} 1.0 \\ .654\,107\,49 \\ .953\,979\,46 \end{pmatrix} = \begin{pmatrix} 5.516\,045\,9 \\ 3.608\,086\,9 \\ 5.262\,194\,5 \end{pmatrix}.$$

13.2.1 Complex Eigenvalues

What about complex eigenvalues? If your matrix is real, you won't see these by graphing the characteristic equation on your calculator. Will the shifted inverse power method find these eigenvalues and their associated eigenvectors? The answer is yes. However, for a real matrix, you must pick α to be complex. This is because the eigenvalues occur in conjugate pairs so if you don't pick it complex, it will be the same distance between any conjugate pair of complex numbers and so nothing in the above argument for convergence implies you will get convergence to a complex number. Also, the process of iteration will yield only real vectors and scalars.

Example 13.2.5 *Find the complex eigenvalues and corresponding eigenvectors for the matrix,*

$$\begin{pmatrix} 5 & -8 & 6 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Here the characteristic equation is $\lambda^3 - 5\lambda^2 + 8\lambda - 6 = 0$. One solution is $\lambda = 3$. The other two are $1+i$ and $1-i$. We will apply the process to $\alpha = i$ so we will find the eigenvalue closest to i.

$$(A - \alpha I)^{-1} = \begin{pmatrix} -.02 - .14i & 1.24 + .68i & -.84 + .12i \\ -.14 + .02i & .68 - .24i & .12 + .84i \\ .02 + .14i & -.24 - .68i & .84 + .88i \end{pmatrix}$$

Then let $\mathbf{u}_1 = (1,1,1)^T$ for lack of any insight into anything better.

$$\begin{pmatrix} -.02 - .14i & 1.24 + .68i & -.84 + .12i \\ -.14 + .02i & .68 - .24i & .12 + .84i \\ .02 + .14i & -.24 - .68i & .84 + .88i \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} .38 + .66i \\ .66 + .62i \\ .62 + .34i \end{pmatrix}$$

$S_2 = .66 + .62i$.

$$\mathbf{u}_2 = \begin{pmatrix} .804\,878\,05 + .243\,902\,44i \\ 1.0 \\ .756\,097\,56 - .195\,121\,95i \end{pmatrix}$$

$$\begin{pmatrix} -.02 - .14i & 1.24 + .68i & -.84 + .12i \\ -.14 + .02i & .68 - .24i & .12 + .84i \\ .02 + .14i & -.24 - .68i & .84 + .88i \end{pmatrix} \cdot$$

$$\begin{pmatrix} .804\,878\,05 + .243\,902\,44i \\ 1.0 \\ .756\,097\,56 - .195\,121\,95i \end{pmatrix}$$

$$= \begin{pmatrix} .646\,341\,46 + .817\,073\,17i \\ .817\,073\,17 + .353\,658\,54i \\ .548\,780\,49 - 6.097\,560\,9 \times 10^{-2} i \end{pmatrix}$$

$S_3 = .646\,341\,46 + .817\,073\,17i$. After more iterations, of this sort, you find $S_9 = 1.002\,748\,5 + 2.137\,621\,7 \times 10^{-4} i$ and

$$\mathbf{u}_9 = \begin{pmatrix} 1.0 \\ .501\,514\,17 - .499\,807\,33i \\ 1.562\,088\,1 \times 10^{-3} - .499\,778\,55i \end{pmatrix}.$$

Then

$$\begin{pmatrix} -.02-.14i & 1.24+.68i & -.84+.12i \\ -.14+.02i & .68-.24i & .12+.84i \\ .02+.14i & -.24-.68i & .84+.88i \end{pmatrix} \cdot \begin{pmatrix} 1.0 \\ .501\,514\,17-.499\,807\,33i \\ 1.562\,088\,1\times10^{-3}-.499\,778\,55i \end{pmatrix} = \begin{pmatrix} 1.000\,407\,8+1.269\,979\times10^{-3}i \\ .501\,077\,31-.498\,893\,66i \\ 8.848\,928\times10^{-4}-.499\,515\,22i \end{pmatrix}$$

$S_{10} = 1.000\,407\,8 + 1.269\,979 \times 10^{-3} i$.

$$\mathbf{u}_{10} = \begin{pmatrix} 1.0 \\ .500\,239\,18-.499\,325\,33i \\ 2.506\,749\,2\times10^{-4}-.499\,311\,92i \end{pmatrix}$$

The scaling factors are not changing much at this point

$$1.000\,407\,8 + 1.269\,979 \times 10^{-3} i = \frac{1}{\lambda - i}$$

The approximate eigenvalue is then $\lambda = .999\,590\,76 + .998\,731\,06i$. This is pretty close to $1+i$. How well does the eigenvector work?

$$\begin{pmatrix} 5 & -8 & 6 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1.0 \\ .500\,239\,18-.499\,325\,33i \\ 2.506\,749\,2\times10^{-4}-.499\,311\,92i \end{pmatrix} = \begin{pmatrix} .999\,590\,61+.998\,731\,12i \\ 1.0 \\ .500\,239\,18-.499\,325\,33i \end{pmatrix}$$

$$(.999\,590\,76+.998\,731\,06i) \begin{pmatrix} 1.0 \\ .500\,239\,18-.499\,325\,33i \\ 2.506\,749\,2\times10^{-4}-.499\,311\,92i \end{pmatrix} = \begin{pmatrix} .999\,590\,76+.998\,731\,06i \\ .998\,726\,18+4.834\,203\,9\times10^{-4}i \\ .498\,928\,9-.498\,857\,22i \end{pmatrix}$$

It took more iterations than before because α was not very close to $1+i$.

This illustrates an interesting topic which leads to many related topics. If you have a polynomial, $x^4 + ax^3 + bx^2 + cx + d$, you can consider it as the characteristic polynomial of a certain matrix, called a **companion matrix**. In this case,

$$\begin{pmatrix} -a & -b & -c & -d \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

The above example was just a companion matrix for $\lambda^3 - 5\lambda^2 + 8\lambda - 6$. You can see the pattern which will enable you to obtain a companion matrix for any polynomial of the form $\lambda^n + a_1\lambda^{n-1} + \cdots + a_{n-1}\lambda + a_n$. This illustrates that one way to find the complex zeros of a polynomial is to use the shifted inverse power method on a companion matrix for the polynomial. Doubtless there are better ways but this does illustrate how impressive this procedure is. Do you have a better way?

13.3 The Rayleigh Quotient

There are many specialized results concerning the eigenvalues and eigenvectors for Hermitian matrices. A matrix, A is Hermitian if $A = A^*$ where A^* means to take the transpose of the conjugate of A. In the case of a real matrix, Hermitian reduces to symmetric. Recall also that for $\mathbf{x} \in \mathbb{F}^n$,

$$|\mathbf{x}|^2 = \mathbf{x}^*\mathbf{x} = \sum_{j=1}^{n} |x_j|^2 .$$

The following corollary gives the theoretical foundation for the spectral theory of Hermitian matrices. This is a corollary of a theorem which is proved Corollary 11.2.14 and Theorem 11.2.13 on Page 217.

Corollary 13.3.1 *If A is Hermitian, then all the eigenvalues of A are real and there exists an orthonormal basis of eigenvectors.*

Thus for $\{\mathbf{x}_k\}_{k=1}^n$ this orthonormal basis,

$$\mathbf{x}_i^*\mathbf{x}_j = \delta_{ij} \equiv \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

For $\mathbf{x} \in \mathbb{F}^n$, $\mathbf{x} \neq \mathbf{0}$, the **Rayleigh quotient** is defined by

$$\frac{\mathbf{x}^* A \mathbf{x}}{|\mathbf{x}|^2}.$$

Now let the eigenvalues of A be $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ and $A\mathbf{x}_k = \lambda_k \mathbf{x}_k$ where $\{\mathbf{x}_k\}_{k=1}^n$ is the above orthonormal basis of eigenvectors mentioned in the corollary. Then if $\mathbf{x}$ is an arbitrary vector, there exist constants, a_i such that

$$\mathbf{x} = \sum_{i=1}^{n} a_i \mathbf{x}_i.$$

Also,

$$\begin{aligned} |\mathbf{x}|^2 &= \sum_{i=1}^{n} \overline{a}_i \mathbf{x}_i^* \sum_{j=1}^{n} a_j \mathbf{x}_j \\ &= \sum_{ij} \overline{a}_i a_j \mathbf{x}_i^* \mathbf{x}_j = \sum_{ij} \overline{a}_i a_j \delta_{ij} = \sum_{i=1}^{n} |a_i|^2 . \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{\mathbf{x}^* A \mathbf{x}}{|\mathbf{x}|^2} &= \frac{\left(\sum_{i=1}^n \overline{a}_i \mathbf{x}_i^*\right)\left(\sum_{j=1}^n a_j \lambda_j \mathbf{x}_j\right)}{\sum_{i=1}^n |a_i|^2} \\ &= \frac{\sum_{ij} \overline{a}_i a_j \lambda_j \mathbf{x}_i^* \mathbf{x}_j}{\sum_{i=1}^n |a_i|^2} = \frac{\sum_{ij} \overline{a}_i a_j \lambda_j \delta_{ij}}{\sum_{i=1}^n |a_i|^2} \\ &= \frac{\sum_{i=1}^n |a_i|^2 \lambda_i}{\sum_{i=1}^n |a_i|^2} \in [\lambda_1, \lambda_n] . \end{aligned}$$

In other words, the Rayleigh quotient is always between the largest and the smallest eigenvalues of A. When $\mathbf{x} = \mathbf{x}_n$, the Rayleigh quotient equals the largest eigenvalue and when $\mathbf{x} = \mathbf{x}_1$ the Rayleigh quotient equals the smallest eigenvalue. Suppose you calculate a Rayleigh quotient. How close is it to some eigenvalue?

Theorem 13.3.2 *Let* $\mathbf{x} \neq \mathbf{0}$ *and form the* ***Rayleigh quotient****,*

$$\frac{\mathbf{x}^* A\mathbf{x}}{|\mathbf{x}|^2} \equiv q.$$

Then there exists an eigenvalue of A*, denoted here by* λ_q *such that*

$$|\lambda_q - q| \leq \frac{|A\mathbf{x} - q\mathbf{x}|}{|\mathbf{x}|}. \tag{13.3}$$

Proof: Let $\mathbf{x} = \sum_{k=1}^n a_k \mathbf{x}_k$ where $\{\mathbf{x}_k\}_{k=1}^n$ is the orthonormal basis of eigenvectors.

$$\begin{aligned}
|A\mathbf{x} - q\mathbf{x}|^2 &= (A\mathbf{x} - q\mathbf{x})^* (A\mathbf{x} - q\mathbf{x}) \\
&= \left(\sum_{k=1}^n a_k \lambda_k \mathbf{x}_k - q a_k \mathbf{x}_k\right)^* \left(\sum_{k=1}^n a_k \lambda_k \mathbf{x}_k - q a_k \mathbf{x}_k\right) \\
&= \left(\sum_{j=1}^n (\lambda_j - q)\, \overline{a}_j \mathbf{x}_j^*\right) \left(\sum_{k=1}^n (\lambda_k - q)\, a_k \mathbf{x}_k\right) \\
&= \sum_{j,k} (\lambda_j - q)\, \overline{a}_j\, (\lambda_k - q)\, a_k \mathbf{x}_j^* \mathbf{x}_k \\
&= \sum_{k=1}^n |a_k|^2 (\lambda_k - q)^2
\end{aligned}$$

Now pick the eigenvalue, λ_q which is closest to q. Then

$$|A\mathbf{x} - q\mathbf{x}|^2 = \sum_{k=1}^n |a_k|^2 (\lambda_k - q)^2 \geq (\lambda_q - q)^2 \sum_{k=1}^n |a_k|^2 = (\lambda_q - q)^2 |\mathbf{x}|^2$$

which implies 13.3.

Example 13.3.3 *Consider the symmetric matrix,* $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \\ 3 & 1 & 4 \end{pmatrix}$*. Let* $\mathbf{x} = (1,1,1)^T$*. How close is the Rayleigh quotient to some eigenvalue of* A*? Find the eigenvector and eigenvalue to several decimal places.*

Everything is real and so there is no need to worry about taking conjugates. Therefore, the Rayleigh quotient is

$$\frac{\begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \\ 3 & 1 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}}{3} = \frac{19}{3}$$

According to the above theorem, there is some eigenvalue of this matrix, λ_q such that

$$\begin{aligned}\left|\lambda_q - \frac{19}{3}\right| &\leq \frac{\left|\begin{pmatrix}1&2&3\\2&2&1\\3&1&4\end{pmatrix}\begin{pmatrix}1\\1\\1\end{pmatrix} - \frac{19}{3}\begin{pmatrix}1\\1\\1\end{pmatrix}\right|}{\sqrt{3}}\\ &= \frac{1}{\sqrt{3}}\begin{pmatrix}-\frac{1}{3}\\-\frac{4}{3}\\\frac{5}{3}\end{pmatrix}\\ &= \frac{\sqrt{\frac{1}{9}+\left(\frac{4}{3}\right)^2+\left(\frac{5}{3}\right)^2}}{\sqrt{3}} = 1.2472\end{aligned}$$

Could you find this eigenvalue and associated eigenvector? Of course you could. This is what the inverse shifted power method is all about.

Solve

$$\left(\begin{pmatrix}1&2&3\\2&2&1\\3&1&4\end{pmatrix} - \frac{19}{3}\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}\right)\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix}$$

In other words solve

$$\begin{pmatrix}-\frac{16}{3}&2&3\\2&-\frac{13}{3}&1\\3&1&-\frac{7}{3}\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix}$$

and divide by the entry which is largest, 3.8707, to get

$$\mathbf{u}_2 = \begin{pmatrix}.69925\\.49389\\1.0\end{pmatrix}$$

Now solve

$$\begin{pmatrix}-\frac{16}{3}&2&3\\2&-\frac{13}{3}&1\\3&1&-\frac{7}{3}\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}.69925\\.49389\\1.0\end{pmatrix}$$

and divide by the entry with largest absolute value, 2.9979 to get

$$\mathbf{u}_3 = \begin{pmatrix}.71473\\.52263\\1.0\end{pmatrix}$$

Now solve

$$\begin{pmatrix}-\frac{16}{3}&2&3\\2&-\frac{13}{3}&1\\3&1&-\frac{7}{3}\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{pmatrix}.71473\\.52263\\1.0\end{pmatrix}$$

and divide by the entry with largest absolute value, 3.0454, to get

$$\mathbf{u}_4 = \begin{pmatrix}.7137\\.52056\\1.0\end{pmatrix}$$

Solve

$$\begin{pmatrix} -\frac{16}{3} & 2 & 3 \\ 2 & -\frac{13}{3} & 1 \\ 3 & 1 & -\frac{7}{3} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} .713\,7 \\ .520\,56 \\ 1.0 \end{pmatrix}$$

and divide by the largest entry, $3.042\,1$ to get

$$\mathbf{u}_5 = \begin{pmatrix} .713\,78 \\ .520\,73 \\ 1.0 \end{pmatrix}$$

You can see these scaling factors are not changing much. The predicted eigenvalue is obtained by solving

$$\frac{1}{\lambda - \frac{19}{3}} = 3.042\,1$$

to obtain $\lambda = 6.6621$. How close is this?

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \\ 3 & 1 & 4 \end{pmatrix} \begin{pmatrix} .713\,78 \\ .520\,73 \\ 1.0 \end{pmatrix} = \begin{pmatrix} 4.755\,2 \\ 3.469 \\ 6.662\,1 \end{pmatrix}$$

while

$$6.662\,1 \begin{pmatrix} .713\,78 \\ .520\,73 \\ 1.0 \end{pmatrix} = \begin{pmatrix} 4.755\,3 \\ 3.469\,2 \\ 6.662\,1 \end{pmatrix}.$$

You see that for practical purposes, this has found the eigenvalue and an eigenvector.

Vector Spaces

It is time to consider the idea of a Vector space.

Definition 14.0.4 *A vector space is an Abelian group of "vectors" satisfying the axioms of an Abelian group,*

$$\mathbf{v}+\mathbf{w}=\mathbf{w}+\mathbf{v},$$

the commutative law of addition,

$$(\mathbf{v}+\mathbf{w})+\mathbf{z}=\mathbf{v}+(\mathbf{w}+\mathbf{z}),$$

the associative law for addition,

$$\mathbf{v}+\mathbf{0}=\mathbf{v},$$

the existence of an additive identity,

$$\mathbf{v}+(-\mathbf{v})=\mathbf{0},$$

the existence of an additive inverse, along with a field of "scalars", $\mathbb{F}$ which are allowed to multiply the vectors according to the following rules. (The Greek letters denote scalars.)

$$\alpha(\mathbf{v}+\mathbf{w})=\alpha\mathbf{v}+\alpha\mathbf{w}, \tag{14.1}$$

$$(\alpha+\beta)\mathbf{v}=\alpha\mathbf{v}+\beta\mathbf{v}, \tag{14.2}$$

$$\alpha(\beta\mathbf{v})=\alpha\beta(\mathbf{v}), \tag{14.3}$$

$$1\mathbf{v}=\mathbf{v}. \tag{14.4}$$

The field of scalars is usually $\mathbb{R}$ or $\mathbb{C}$ and the vector space will be called real or complex depending on whether the field is $\mathbb{R}$ or $\mathbb{C}$. However, other fields are also possible. For example, one could use the field of rational numbers or even the field of the integers mod p for p a prime. A vector space is also called a linear space.

For example, $\mathbb{R}^n$ with the usual conventions is an example of a real vector space and $\mathbb{C}^n$ is an example of a complex vector space. Up to now, the discussion has been for $\mathbb{R}^n$ or $\mathbb{C}^n$ and all that is taking place is an increase in generality and abstraction.

Definition 14.0.5 *If $\{\mathbf{v}_1,\cdots,\mathbf{v}_n\}\subseteq V$, a vector space, then*

$$\mathrm{span}(\mathbf{v}_1,\cdots,\mathbf{v}_n)\equiv\left\{\sum_{i=1}^{n}\alpha_i\mathbf{v}_i:\alpha_i\in\mathbb{F}\right\}.$$

A subset, $W\subseteq V$ is said to be a subspace if it is also a vector space with the same field of scalars. Thus $W\subseteq V$ is a subspace if $ax+by\in W$ whenever $a,b\in\mathbb{F}$ and $x,y\in W$. The span of a set of vectors as just described is an example of a subspace.

Definition 14.0.6 *If* $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\} \subseteq V$, *the set of vectors is linearly independent if*

$$\sum_{i=1}^{n} \alpha_i \mathbf{v}_i = \mathbf{0}$$

implies

$$\alpha_1 = \cdots = \alpha_n = 0$$

and $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ *is called a basis for* V *if*

$$\text{span}(\mathbf{v}_1, \cdots, \mathbf{v}_n) = V$$

and $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ *is linearly independent. The set of vectors is linearly dependent if it is not linearly independent.*

The next theorem is called the exchange theorem. It is very important that you understand this theorem. There are two kinds of people who go further in linear algebra, those who understand this theorem and its corollary presented later and those who don't. Those who do understand these theorems are able to proceed and learn more linear algebra while those who don't are doomed to wander in the wilderness of confusion and sink into the swamp of despair. Therefore, I am giving multiple proofs. Try to understand at least one of them. Several amount to the same thing, just worded differently.

Theorem 14.0.7 *Let* $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ *be a linearly independent set of vectors such that each* $\mathbf{x}_i$ *is in the span*$\{\mathbf{y}_1, \cdots, \mathbf{y}_s\}$. *Then* $r \leq s$.

Proof 1: Let

$$\mathbf{x}_k = \sum_{j=1}^{s} a_{jk} \mathbf{y}_j$$

If $r > s$, then the matrix $A = (a_{jk})$ has more columns than rows. By Corollary 7.2.8 one of these columns is a linear combination of the others. This implies there exist scalars $c_1, \cdots, c_r$ such that

$$\sum_{k=1}^{r} a_{jk} c_k = 0, \ j = 1, \cdots, r$$

Then

$$\sum_{k=1}^{r} c_k \mathbf{x}_k = \sum_{k=1}^{r} c_k \sum_{j=1}^{s} a_{jk} \mathbf{y}_j = \sum_{j=1}^{s} \left(\sum_{k=1}^{r} c_k a_{jk} \right) \mathbf{y}_j = \mathbf{0}$$

which contradicts the assumption that $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ is linearly independent. Hence $r \leq s$.

Proof 2: Define $\text{span}\{\mathbf{y}_1, \cdots, \mathbf{y}_s\} \equiv V$, it follows there exist scalars, $c_1, \cdots, c_s$ such that

$$\mathbf{x}_1 = \sum_{i=1}^{s} c_i \mathbf{y}_i. \tag{14.5}$$

Not all of these scalars can equal zero because if this were the case, it would follow that $\mathbf{x}_1 = 0$ and so $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ would not be linearly independent. Indeed, if $\mathbf{x}_1 = 0$, $1\mathbf{x}_1 + \sum_{i=2}^{r} 0\mathbf{x}_i = \mathbf{x}_1 = 0$ and so there would exist a nontrivial linear combination of the vectors $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ which equals zero.

Say $c_k \neq 0$. Then solve (14.5) for $\mathbf{y}_k$ and obtain

$$\mathbf{y}_k \in \text{span}\left(\mathbf{x}_1, \overbrace{\mathbf{y}_1, \cdots, \mathbf{y}_{k-1}, \mathbf{y}_{k+1}, \cdots, \mathbf{y}_s}^{\text{s-1 vectors here}} \right).$$

Define $\{\mathbf{z}_1, \cdots, \mathbf{z}_{s-1}\}$ by

$$\{\mathbf{z}_1, \cdots, \mathbf{z}_{s-1}\} \equiv \{\mathbf{y}_1, \cdots, \mathbf{y}_{k-1}, \mathbf{y}_{k+1}, \cdots, \mathbf{y}_s\}$$

Therefore, $\text{span}\{\mathbf{x}_1, \mathbf{z}_1, \cdots, \mathbf{z}_{s-1}\} = V$ because if $\mathbf{v} \in V$, there exist constants $c_1, \cdots, c_s$ such that

$$\mathbf{v} = \sum_{i=1}^{s-1} c_i \mathbf{z}_i + c_s \mathbf{y}_k.$$

Now replace the $\mathbf{y}_k$ in the above with a linear combination of the vectors, $\{\mathbf{x}_1, \mathbf{z}_1, \cdots, \mathbf{z}_{s-1}\}$ to obtain $\mathbf{v} \in \text{span}\{\mathbf{x}_1, \mathbf{z}_1, \cdots, \mathbf{z}_{s-1}\}$. The vector $\mathbf{y}_k$, in the list $\{\mathbf{y}_1, \cdots, \mathbf{y}_s\}$, has now been replaced with the vector $\mathbf{x}_1$ and the resulting modified list of vectors has the same span as the original list of vectors, $\{\mathbf{y}_1, \cdots, \mathbf{y}_s\}$.

Now suppose that $r > s$ and that $\text{span}\{\mathbf{x}_1, \cdots, \mathbf{x}_l, \mathbf{z}_1, \cdots, \mathbf{z}_p\} = V$ where the vectors, $\mathbf{z}_1, \cdots, \mathbf{z}_p$ are each taken from the set, $\{\mathbf{y}_1, \cdots, \mathbf{y}_s\}$ and $l + p = s$. This has now been done for $l = 1$ above. Then since $r > s$, it follows that $l \leq s < r$ and so $l + 1 \leq r$. Therefore, $\mathbf{x}_{l+1}$ is a vector not in the list, $\{\mathbf{x}_1, \cdots, \mathbf{x}_l\}$ and since $\text{span}\{\mathbf{x}_1, \cdots, \mathbf{x}_l, \mathbf{z}_1, \cdots, \mathbf{z}_p\} = V$ there exist scalars, c_i and d_j such that

$$\mathbf{x}_{l+1} = \sum_{i=1}^{l} c_i \mathbf{x}_i + \sum_{j=1}^{p} d_j \mathbf{z}_j. \tag{14.6}$$

Now not all the d_j can equal zero because if this were so, it would follow that $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ would be a linearly dependent set because one of the vectors would equal a linear combination of the others. Therefore, (14.6) can be solved for one of the $\mathbf{z}_i$, say $\mathbf{z}_k$, in terms of $\mathbf{x}_{l+1}$ and the other $\mathbf{z}_i$ and just as in the above argument, replace that $\mathbf{z}_i$ with $\mathbf{x}_{l+1}$ to obtain

$$\text{span}\left(\mathbf{x}_1, \cdots \mathbf{x}_l, \mathbf{x}_{l+1}, \overbrace{\mathbf{z}_1, \cdots \mathbf{z}_{k-1}, \mathbf{z}_{k+1}, \cdots, \mathbf{z}_p}^{\text{p-1 vectors here}}\right) = V.$$

Continue this way, eventually obtaining

$$\text{span}(\mathbf{x}_1, \cdots, \mathbf{x}_s) = V.$$

But then $\mathbf{x}_r \in \text{span}\{\mathbf{x}_1, \cdots, \mathbf{x}_s\}$ contrary to the assumption that $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ is linearly independent. Therefore, $r \leq s$ as claimed.

Proof 3: Let $V \equiv \text{span}(\mathbf{y}_1, \cdots, \mathbf{y}_s)$ and suppose $r > s$. Let $A_l \equiv \{\mathbf{x}_1, \cdots, \mathbf{x}_l\}, A_0 = \emptyset$, and let B_{s-l} denote a subset of the vectors, $\{\mathbf{y}_1, \cdots, \mathbf{y}_s\}$ which contains $s - l$ vectors and has the property that $\text{span}(A_l, B_{s-l}) = V$. Note that the assumption of the theorem says $\text{span}(A_0, B_s) = V$.

Now an exchange operation is given for $\text{span}(A_l, B_{s-l}) = V$. Since $r > s$, it follows $l < r$. Letting

$$B_{s-l} \equiv \{\mathbf{z}_1, \cdots, \mathbf{z}_{s-l}\} \subseteq \{\mathbf{y}_1, \cdots, \mathbf{y}_s\},$$

it follows there exist constants, c_i and d_i such that

$$\mathbf{x}_{l+1} = \sum_{i=1}^{l} c_i \mathbf{x}_i + \sum_{i=1}^{s-l} d_i \mathbf{z}_i,$$

and not all the d_i can equal zero. (If they were all equal to zero, it would follow that the set, $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ would be dependent since one of the vectors in it would be a linear combination of the others.)

Let $d_k \neq 0$. Then $\mathbf{z}_k$ can be solved for as follows.

$$\mathbf{z}_k = \frac{1}{d_k}\mathbf{x}_{l+1} - \sum_{i=1}^{l} \frac{c_i}{d_k}\mathbf{x}_i - \sum_{i \neq k} \frac{d_i}{d_k}\mathbf{z}_i.$$

This implies $V = \text{span}\,(A_{l+1}, B_{s-l-1})$, where $B_{s-l-1} \equiv B_{s-l} \setminus \{\mathbf{z}_k\}$, a set obtained by deleting $\mathbf{z}_k$ from B_{k-l}. You see, the process exchanged a vector in B_{s-l} with one from $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ and kept the span the same. Starting with $V = \text{span}\,(A_0, B_s)$, do the exchange operation until $V = \text{span}\,(A_{s-1}, \mathbf{z})$ where $\mathbf{z} \in \{\mathbf{y}_1, \cdots, \mathbf{y}_s\}$. Then one more application of the exchange operation yields $V = \text{span}\,(A_s)$. But this implies $\mathbf{x}_r \in \text{span}\,(A_s) = \text{span}\,(\mathbf{x}_1, \cdots, \mathbf{x}_s)$, contradicting the linear independence of $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$. It follows that $r \leq s$ as claimed.

Proof 4: Suppose $r > s$. Let $\mathbf{z}_k$ denote a vector of $\{\mathbf{y}_1, \cdots, \mathbf{y}_s\}$. Thus there exists j as small as possible such that

$$\text{span}\,(\mathbf{y}_1, \cdots, \mathbf{y}_s) = \text{span}\,(\mathbf{x}_1, \cdots, \mathbf{x}_m, \mathbf{z}_1, \cdots, \mathbf{z}_j)$$

where $m + j = s$. It is given that $m = 0$, corresponding to no vectors of $\{\mathbf{x}_1, \cdots, \mathbf{x}_m\}$ and $j = s$, corresponding to all the $\mathbf{y}_k$ results in the above equation holding. If $j > 0$ then $m < s$ and so

$$\mathbf{x}_{m+1} = \sum_{k=1}^{m} a_k \mathbf{x}_k + \sum_{i=1}^{j} b_i \mathbf{z}_i$$

Not all the b_i can equal 0 and so you can solve for one of them in terms of $\mathbf{x}_{m+1}, \mathbf{x}_m, \cdots, \mathbf{x}_1$, and the other $\mathbf{z}_k$. Therefore, there exists

$$\{\mathbf{z}_1, \cdots, \mathbf{z}_{j-1}\} \subseteq \{\mathbf{y}_1, \cdots, \mathbf{y}_s\}$$

such that

$$\text{span}\,(\mathbf{y}_1, \cdots, \mathbf{y}_s) = \text{span}\,(\mathbf{x}_1, \cdots, \mathbf{x}_{m+1}, \mathbf{z}_1, \cdots, \mathbf{z}_{j-1})$$

contradicting the choice of j. Hence $j = 0$ and

$$\text{span}\,(\mathbf{y}_1, \cdots, \mathbf{y}_s) = \text{span}\,(\mathbf{x}_1, \cdots, \mathbf{x}_s)$$

It follows that

$$\mathbf{x}_{s+1} \in \text{span}\,(\mathbf{x}_1, \cdots, \mathbf{x}_s)$$

contrary to the assumption the $\mathbf{x}_k$ are linearly independent. Therefore, $r \leq s$ as claimed. This proves the theorem.

Corollary 14.0.8 *If $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ and $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ are two bases for V, then $m = n$.*

Proof: By Theorem 14.0.7, $m \leq n$ and $n \leq m$.

This corollary is very important so here is another proof of it given independent of the exchange theorem above.

Theorem 14.0.9 *Let V be a vector space and suppose $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ and $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ are two bases for V. Then $k = m$.*

Proof: Suppose $k > m$. Then since the vectors, $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ span V, there exist scalars, c_{ij} such that

$$\sum_{i=1}^{m} c_{ij}\mathbf{v}_i = \mathbf{u}_j.$$

Therefore,

$$\sum_{j=1}^{k} d_j \mathbf{u}_j = \mathbf{0} \text{ if and only if } \sum_{j=1}^{k}\sum_{i=1}^{m} c_{ij} d_j \mathbf{v}_i = \mathbf{0}$$

if and only if

$$\sum_{i=1}^{m} \left(\sum_{j=1}^{k} c_{ij} d_j \right) \mathbf{v}_i = \mathbf{0}$$

Now since$\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is independent, this happens if and only if

$$\sum_{j=1}^{k} c_{ij} d_j = 0, \ i = 1, 2, \cdots, m.$$

However, this is a system of m equations in k variables, $d_1, \cdots, d_k$ and $m < k$. Therefore, there exists a solution to this system of equations in which not all the d_j are equal to zero. Recall why this is so. The augmented matrix for the system is of the form $\begin{pmatrix} C & \mathbf{0} \end{pmatrix}$ where C is a matrix which has more columns than rows. Therefore, there are free variables and hence nonzero solutions to the system of equations. However, this contradicts the linear independence of $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ because, as explained above, $\sum_{j=1}^{k} d_j \mathbf{u}_j = \mathbf{0}$. Similarly it cannot happen that $m > k$. This proves the theorem.

Definition 14.0.10 *A vector space V is of dimension n if it has a basis consisting of n vectors. This is well defined thanks to Corollary 14.0.8. It is always assumed here that $n < \infty$ in this case, such a vector space is said to be finite dimensional.*

Theorem 14.0.11 *If $V = \text{span}(\mathbf{u}_1, \cdots, \mathbf{u}_n)$ then some subset of $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ is a basis for V. Also, if $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\} \subseteq V$ is linearly independent and the vector space is finite dimensional, then the set, $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$, can be enlarged to obtain a basis of V.*

Proof: Let

$$S = \{E \subseteq \{\mathbf{u}_1, \cdots, \mathbf{u}_n\} \text{ such that } \text{span}(E) = V\}.$$

For $E \in S$, let $|E|$ denote the number of elements of E. Let

$$m \equiv \min\{|E| \text{ such that } E \in S\}.$$

Thus there exist vectors

$$\{\mathbf{v}_1, \cdots, \mathbf{v}_m\} \subseteq \{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$$

such that

$$\text{span}(\mathbf{v}_1, \cdots, \mathbf{v}_m) = V$$

and m is as small as possible for this to happen. If this set is linearly independent, it follows it is a basis for V and the theorem is proved. On the other hand, if the set is not linearly independent, then there exist scalars,

$$c_1, \cdots, c_m$$

such that

$$\mathbf{0} = \sum_{i=1}^{m} c_i \mathbf{v}_i$$

and not all the c_i are equal to zero. Suppose $c_k \neq 0$. Then the vector, $\mathbf{v}_k$ may be solved for in terms of the other vectors. Consequently,

$$V = \text{span}(\mathbf{v}_1, \cdots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \cdots, \mathbf{v}_m)$$

contradicting the definition of m. This proves the first part of the theorem.

To obtain the second part, begin with $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ and suppose a basis for V is $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$. If

$$\text{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k) = V,$$

then $k = n$. If not, there exists a vector,

$$\mathbf{u}_{k+1} \notin \text{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k).$$

Then $\{\mathbf{u}_1, \cdots, \mathbf{u}_k, \mathbf{u}_{k+1}\}$ is also linearly independent. Continue adding vectors in this way until n linearly independent vectors have been obtained. Then $\text{span}(\mathbf{u}_1, \cdots, \mathbf{u}_n) = V$ because if it did not do so, there would exist $\mathbf{u}_{n+1}$ as just described and $\{\mathbf{u}_1, \cdots, \mathbf{u}_{n+1}\}$ would be a linearly independent set of vectors having $n+1$ elements even though $\{\mathbf{v}_1, \cdots, v_n\}$ is a basis. This would contradict Theorem 14.0.7. Therefore, this list is a basis and this proves the theorem.

It is useful to emphasize some of the ideas used in the above proof.

Lemma 14.0.12 *Suppose* $\mathbf{v} \notin \text{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k)$ *and* $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ *is linearly independent. Then* $\{\mathbf{u}_1, \cdots, \mathbf{u}_k, \mathbf{v}\}$ *is also linearly independent.*

Proof: Suppose $\sum_{i=1}^{k} c_i \mathbf{u}_i + d\mathbf{v} = 0$. It is required to verify that each $c_i = 0$ and that $d = 0$. But if $d \neq 0$, then you can solve for $\mathbf{v}$ as a linear combination of the vectors, $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$,

$$\mathbf{v} = -\sum_{i=1}^{k} \left(\frac{c_i}{d}\right) \mathbf{u}_i$$

contrary to assumption. Therefore, $d = 0$. But then $\sum_{i=1}^{k} c_i \mathbf{u}_i = 0$ and the linear independence of $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ implies each $c_i = 0$ also. This proves the lemma.

Theorem 14.0.13 *Let* V *be a nonzero subspace of a finite dimensional vector space,* W *of dimension,* n*. Then* V *has a basis with no more than* n *vectors.*

Proof: Let $\mathbf{v}_1 \in V$ where $\mathbf{v}_1 \neq 0$. If $\text{span}\{\mathbf{v}_1\} = V$, stop. $\{\mathbf{v}_1\}$ is a basis for V. Otherwise, there exists $\mathbf{v}_2 \in V$ which is not in $\text{span}\{\mathbf{v}_1\}$. By Lemma 14.0.12 $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a linearly independent set of vectors. If $\text{span}\{\mathbf{v}_1, \mathbf{v}_2\} = V$ stop, $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for V. If $\text{span}\{\mathbf{v}_1, \mathbf{v}_2\} \neq V$, then there exists $\mathbf{v}_3 \notin \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$ and $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a larger linearly independent set of vectors. Continuing this way, the process must stop before $n + 1$ steps because if not, it would be possible to obtain $n + 1$ linearly independent vectors contrary to the exchange theorem, Theorems 14.0.7. This proves the theorem.

Linear Transformations

15.1 Matrix Multiplication As A Linear Transformation

Definition 15.1.1 *Let V and W be two finite dimensional vector spaces. A function, L which maps V to W is called a linear transformation and $L \in \mathcal{L}(V,W)$ if for all scalars α and β, and vectors $\mathbf{v}, \mathbf{w}$,*

$$L(\alpha\mathbf{v}+\beta\mathbf{w}) = \alpha L(\mathbf{v}) + \beta L(\mathbf{w}).$$

An example of a linear transformation is familiar matrix multiplication. Let $A = (a_{ij})$ be an $m \times n$ matrix. Then an example of a linear transformation $L : \mathbb{F}^n \to \mathbb{F}^m$ is given by

$$(L\mathbf{v})_i \equiv \sum_{j=1}^{n} a_{ij} v_j.$$

Here

$$\mathbf{v} \equiv \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{F}^n.$$

15.2 $\mathcal{L}(V,W)$ As A Vector Space

Definition 15.2.1 *Given $L, M \in \mathcal{L}(V,W)$ define a new element of $\mathcal{L}(V,W)$, denoted by $L+M$ according to the rule*

$$(L+M)\mathbf{v} \equiv L\mathbf{v} + M\mathbf{v}.$$

For α a scalar and $L \in \mathcal{L}(V,W)$, define $\alpha L \in \mathcal{L}(V,W)$ by

$$\alpha L(\mathbf{v}) \equiv \alpha(L\mathbf{v}).$$

You should verify that all the axioms of a vector space hold for $\mathcal{L}(V,W)$ with the above definitions of vector addition and scalar multiplication. What about the dimension of $\mathcal{L}(V,W)$?

Theorem 15.2.2 *Let V and W be finite dimensional normed linear spaces of dimension n and m respectively Then* $\dim(\mathcal{L}(V,W)) = mn$.

Proof: Let the two sets of bases be

$$\{\mathbf{v}_1, \cdots, \mathbf{v}_n\} \text{ and } \{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$$

for X and Y respectively. Let $E_{ik} \in \mathcal{L}(V,W)$ be the linear transformation defined on the basis, $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$, by

$$E_{ik}\mathbf{v}_j \equiv \mathbf{w}_i \delta_{jk}$$

where $\delta_{ik} = 1$ if $i = k$ and 0 if $i \neq k$. Then let $L \in \mathcal{L}(V,W)$. Since $\{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ is a basis, there exist constants, d_{jk} such that

$$L\mathbf{v}_r = \sum_{j=1}^{m} d_{jr}\mathbf{w}_j$$

Also

$$\sum_{j=1}^{m}\sum_{k=1}^{n} d_{jk}E_{jk}(\mathbf{v}_r) = \sum_{j=1}^{m} d_{jr}\mathbf{w}_j.$$

It follows that

$$L = \sum_{j=1}^{m}\sum_{k=1}^{n} d_{jk}E_{jk}$$

because the two linear transformations agree on a basis. Since L is arbitrary this shows

$$\{E_{ik} : i = 1, \cdots, m,\ k = 1, \cdots, n\}$$

spans $\mathcal{L}(V,W)$.

If

$$\sum_{i,k} d_{ik}E_{ik} = \mathbf{0},$$

then

$$\mathbf{0} = \sum_{i,k} d_{ik}E_{ik}(\mathbf{v}_l) = \sum_{i=1}^{m} d_{il}\mathbf{w}_i$$

and so, since $\{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ is a basis, $d_{il} = 0$ for each $i = 1, \cdots, m$. Since l is arbitrary, this shows $d_{il} = 0$ for all i and l. Thus these linear transformations form a basis and this shows the dimension of $\mathcal{L}(V,W)$ is mn as claimed.

15.3 Eigenvalues And Eigenvectors Of Linear Transformations

Let V be a finite dimensional vector space. For example, it could be a subspace of $\mathbb{C}^n$. Also suppose $A \in \mathcal{L}(V,V)$. Does A have eigenvalues and eigenvectors just like the case where A is a $n \times n$ matrix?

Theorem 15.3.1 *Let V be a nonzero finite dimensional complex vector space of dimension n. Suppose also the field of scalars equals $\mathbb{C}$.*[1] *Suppose $A \in \mathcal{L}(V,V)$. Then there exists $v \neq 0$ and $\lambda \in \mathbb{C}$ such that*

$$Av = \lambda v.$$

[1] All that is really needed is that the minimal polynomial can be completely factored in the given field. The complex numbers have this property from the fundamental theorem of algebra.

Proof: Consider the linear transformations, $I, A, A^2, \cdots, A^{n^2}$. There are n^2+1 of these transformations and so by Theorem 15.2.2 the set is linearly dependent. Thus there exist constants, $c_i \in \mathbb{C}$ such that

$$c_0 I + \sum_{k=1}^{n^2} c_k A^k = 0.$$

This implies there exists a polynomial, $q(\lambda)$ which has the property that $q(A) = 0$. In fact, $q(\lambda) \equiv c_0 + \sum_{k=1}^{n^2} c_k \lambda^k$. Dividing by the leading term, it can be assumed this polynomial is of the form $\lambda^m + c_{m-1}\lambda^{m-1} + \cdots + c_1\lambda + c_0$, a monic polynomial. Now consider all such monic polynomials, q such that $q(A) = 0$ and pick one which has the smallest degree. This is called the minimal polynomial and will be denoted here by $p(\lambda)$. By the fundamental theorem of algebra, $p(\lambda)$ is of the form

$$p(\lambda) = \prod_{k=1}^{p} (\lambda - \lambda_k).$$

Thus, since p has minimal degree,

$$\prod_{k=1}^{p} (A - \lambda_k I) = 0, \text{ but } \prod_{k=1}^{p-1} (A - \lambda_k I) \neq 0.$$

Therefore, there exists $u \neq 0$ such that

$$v \equiv \left(\prod_{k=1}^{p-1} (A - \lambda_k I) \right)(u) \neq 0.$$

But then

$$(A - \lambda_p I)\, v = (A - \lambda_p I) \left(\prod_{k=1}^{p-1} (A - \lambda_k I) \right)(u) = 0.$$

This proves the theorem.

Corollary 15.3.2 *In the above theorem, each of the scalars, λ_k has the property that there exists a nonzero v such that $(A - \lambda_i I)\, v = 0$. Furthermore the λ_i are the only scalars with this property.*

Proof: For the first claim, just factor out $(A - \lambda_i I)$ instead of $(A - \lambda_p I)$. Next suppose $(A - \mu I)\, v = 0$ for some μ and $v \neq 0$. Then

$$\begin{aligned}
0 &= \prod_{k=1}^{p} (A - \lambda_k I)\, v = \prod_{k=1}^{p-1} (A - \lambda_k I)\, (Av - \lambda_p v) \\
&= (\mu - \lambda_p) \left(\prod_{k=1}^{p-1} (A - \lambda_k I) \right) v \\
&= (\mu - \lambda_p) \left(\prod_{k=1}^{p-2} (A - \lambda_k I) \right) (Av - \lambda_{p-1} v) \\
&= (\mu - \lambda_p)\, (\mu - \lambda_{p-1}) \left(\prod_{k=1}^{p-2} (A - \lambda_k I) \right)
\end{aligned}$$

continuing this way yields

$$= \prod_{k=1}^{p} (\mu - \lambda_k)\, v,$$

a contradiction unless $\mu = \lambda_k$ for some k.

Therefore, these are eigenvectors and eigenvalues with the usual meaning. This leads to the following definition.

Definition 15.3.3 *For $A \in \mathcal{L}(V,V)$ where $\dim(V) = n$, the scalars, λ_k in the minimal polynomial,*

$$p(\lambda) = \prod_{k=1}^{p} (\lambda - \lambda_k)$$

are called the eigenvalues of A. The collection of eigenvalues of A is denoted by $\sigma(A)$. For λ an eigenvalue of $A \in \mathcal{L}(V,V)$, the generalized eigenspace is defined as

$$V_\lambda \equiv \{x \in V : (A - \lambda I)^m x = 0 \text{ for some } m \in \mathbb{N}\}$$

and the eigenspace is defined as

$$\{x \in V : (A - \lambda I)\, x = 0\} \equiv \ker(A - \lambda I).$$

Also, for subspaces of V, $V_1, V_2, \cdots, V_r$, the symbol, $V_1 + V_2 + \cdots + V_r$ or the shortened version, $\sum_{i=1}^{r} V_i$ will denote the set of all vectors of the form $\sum_{i=1}^{r} v_i$ where $v_i \in V_i$.

Lemma 15.3.4 *The generalized eigenspace for $\lambda \in \sigma(A)$ where $A \in \mathcal{L}(V,V)$ for V an n dimensional vector space is a subspace, V_λ of V satisfying*

$$A : V_\lambda \to V_\lambda,$$

and there exists a smallest integer, m with the property that

$$\ker(A - \lambda I)^m = \left\{x \in V : (A - \lambda I)^k x = 0 \text{ for some } k \in \mathbb{N}\right\}. \tag{15.1}$$

Proof: The claim that the generalized eigenspace is a subspace is obvious. To establish the second part, note that

$$\left\{\ker(A - \lambda I)^k\right\}$$

yields an increasing sequence of subspaces. Eventually

$$\dim(\ker(A - \lambda I)^m) = \dim\left(\ker(A - \lambda I)^{m+1}\right)$$

and so $\ker(A - \lambda I)^m = \ker(A - \lambda I)^{m+1}$. Now if $\mathbf{x} \in \ker(A - \lambda I)^{m+2}$, then

$$(A - \lambda I)\,\mathbf{x} \in \ker(A - \lambda I)^{m+1} = \ker(A - \lambda I)^m$$

and so there exists $\mathbf{y} \in \ker(A - \lambda I)^m$ such that $(A - \lambda I)\,\mathbf{x} = \mathbf{y}$ and consequently

$$(A - \lambda I)^{m+1}\,\mathbf{x} = (A - \lambda I)^m\,\mathbf{y} = \mathbf{0}$$

showing that $\mathbf{x} \in \ker(A - \lambda I)^{m+1}$. Therefore, continuing this way, it follows that for all $k \in \mathbb{N}$,

$$\ker(A - \lambda I)^m = \ker(A - \lambda I)^{m+k}.$$

Therefore, this shows 15.1.

The following theorem is of major importance and will be the basis for the very important theorems concerning block diagonal matrices.

Theorem 15.3.5 *Let V be a complex vector space of dimension n and suppose $\sigma(A) = \{\lambda_1, \cdots, \lambda_k\}$ where the λ_i are the distinct eigenvalues of A. Denote by V_i the generalized eigenspace for λ_i and let r_i be the multiplicity of λ_i. By this is meant that*

$$V_i = \ker (A - \lambda_i I)^{r_i} \tag{15.2}$$

and r_i is the smallest integer with this property. Then

$$V = \sum_{i=1}^{k} V_i. \tag{15.3}$$

Proof: This is proved by induction on k. First suppose there is only one eigenvalue, λ_1 of algebraic multiplicity m. Then by the definition of eigenvalues given in Definition 15.3.3, A satisfies an equation of the form

$$(A - \lambda_1 I)^r = 0$$

where r is as small as possible for this to take place. Thus $\ker (A - \lambda_1 I)^r = V$ and the theorem is proved in the case of one eigenvalue.

Now suppose the theorem is true for any $i \leq k-1$ where $k \geq 2$ and suppose $\sigma(A) = \{\lambda_1, \cdots, \lambda_k\}$.

Claim 1: Let $\mu \neq \lambda_i$, Then $(A - \mu I)^m : V_i \to V_i$ and is one to one and onto for every $m \in \mathbb{N}$.

Proof: It is clear that $(A - \mu I)^m$ maps V_i to V_i because if $v \in V_i$ then $(A - \lambda_i I)^k v = 0$ for some $k \in \mathbb{N}$. Consequently,

$$(A - \lambda_i I)^k (A - \mu I)^m v = (A - \mu I)^m (A - \lambda_i I)^k v = (A - \mu I)^m 0 = 0$$

which shows that $(A - \mu I)^m v \in V_i$.

It remains to verify that $(A - \mu I)^m$ is one to one. This will be done by showing that $(A - \mu I)$ is one to one. Let $w \in V_i$ and suppose $(A - \mu I) w = 0$ so that $Aw = \mu w$. Then for some $m \in \mathbb{N}$, $(A - \lambda_i I)^m w = 0$ and so by the binomial theorem,

$$(\mu - \lambda_i)^m w = \sum_{l=0}^{m} \binom{m}{l} (-\lambda_i)^{m-l} \mu^l w$$

$$\sum_{l=0}^{m} \binom{m}{l} (-\lambda_i)^{m-l} A^l w = (A - \lambda_i I)^m w = 0.$$

Therefore, since $\mu \neq \lambda_i$, it follows $w = 0$ and this verifies $(A - \mu I)$ is one to one. Thus $(A - \mu I)^m$ is also one to one on V_i Letting $\left\{u_1^i, \cdots, u_{r_k}^i\right\}$ be a basis for V_i, it follows $\left\{(A - \mu I)^m u_1^i, \cdots, (A - \mu I)^m u_{r_k}^i\right\}$ is also a basis and so $(A - \mu I)^m$ is also onto.

Let p be the smallest integer such that $\ker (A - \lambda_k I)^p = V_k$ and define

$$W \equiv (A - \lambda_k I)^p (V).$$

Claim 2: $A : W \to W$ and λ_k is not an eigenvalue for A restricted to W.

Proof: Suppose to the contrary that

$$A (A - \lambda_k I)^p u = \lambda_k (A - \lambda_k I)^p u$$

where $(A - \lambda_k I)^p u \neq 0$. Then subtracting $\lambda_k (A - \lambda_k I)^p u$ from both sides yields

$$(A - \lambda_k I)^{p+1} u = 0$$

and so $u \in \ker((A-\lambda_k I)^p)$ from the definition of p. But this requires $(A-\lambda_k I)^p u = 0$ contrary to $(A-\lambda_k I)^p u \neq 0$. This has verified the claim.

It follows from this claim that the eigenvalues of A restricted to W are a subset of $\{\lambda_1, \cdots, \lambda_{k-1}\}$. Letting

$$V_i' \equiv \left\{ w \in W : (A-\lambda_i)^l w = 0 \text{ for some } l \in \mathbb{N} \right\},$$

it follows from the induction hypothesis that

$$W = \sum_{i=1}^{k-1} V_i' \subseteq \sum_{i=1}^{k-1} V_i.$$

From Claim 1, $(A-\lambda_k I)^p$ maps V_i one to one and onto V_i. From the definition of W, if $x \in V$, then $(A-\lambda_k I)^p x \in W$. It follows there exist $x_i \in V_i$ such that

$$(A-\lambda_k I)^p x = \sum_{i=1}^{k-1} \overbrace{(A-\lambda_k I)^p x_i}^{\in V_i}.$$

Consequently

$$(A-\lambda_k I)^p \left(\overbrace{x - \sum_{i=1}^{k-1} x_i}^{\in V_k} \right) = 0$$

and so there exists $x_k \in V_k$ such that

$$x - \sum_{i=1}^{k-1} x_i = x_k$$

and this proves the theorem.

Definition 15.3.6 *Let $\{V_i\}_{i=1}^r$ be subspaces of V which have the property that if $v_i \in V_i$ and*

$$\sum_{i=1}^{r} v_i = 0, \tag{15.4}$$

then $v_i = 0$ for each i. Under this condition, a special notation is used to denote $\sum_{i=1}^r V_i$. This notation is

$$V_1 \oplus \cdots \oplus V_r$$

and it is called a direct sum of subspaces.

Theorem 15.3.7 *Let $\{V_i\}_{i=1}^m$ be subspaces of V which have the property 15.4 and let $B_i = \{u_1^i, \cdots, u_{r_i}^i\}$ be a basis for V_i. Then $\{B_1, \cdots, B_m\}$ is a basis for $V_1 \oplus \cdots \oplus V_m = \sum_{i=1}^m V_i$.*

Proof: It is clear that $\text{span}(B_1, \cdots, B_m) = V_1 \oplus \cdots \oplus V_m$. It only remains to verify that $\{B_1, \cdots, B_m\}$ is linearly independent. Arbitrary elements of $\text{span}(B_1, \cdots, B_m)$ are of the form

$$\sum_{k=1}^{m} \sum_{i=1}^{r_i} c_i^k u_i^k.$$

Suppose then that

$$\sum_{k=1}^{m}\sum_{i=1}^{r_i} c_i^k u_i^k = 0.$$

Since $\sum_{i=1}^{r_i} c_i^k u_i^k \in V_k$ it follows $\sum_{i=1}^{r_i} c_i^k u_i^k = 0$ for each k. But then $c_i^k = 0$ for each $i = 1, \cdots, r_i$. This proves the theorem.

The following corollary is the main result.

Corollary 15.3.8 *Let V be a complex vector space of dimension, n and let $A \in \mathcal{L}(V,V)$. Also suppose $\sigma(A) = \{\lambda_1, \cdots, \lambda_s\}$ where the λ_i are distinct. Then letting V_{λ_i} denote the generalized eigenspace for λ_i,*

$$V = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_s}$$

and if B_i is a basis for V_{λ_i}, then $\{B_1, B_2, \cdots, B_s\}$ is a basis for V.

Proof: It is necessary to verify that the V_{λ_i} satisfy condition 15.4. Let

$$V_{\lambda_i} = \ker(A - \lambda_i I)^{r_i}$$

and suppose $v_i \in V_{\lambda_i}$ and $\sum_{i=1}^{k} v_i = 0$ where $k \leq s$. It is desired to show this implies each $v_i = 0$. It is clearly true if $k = 1$. Suppose then that the condition holds for $k-1$ and

$$\sum_{i=1}^{k} v_i = 0$$

and not all the $v_i = 0$. By Claim 1 in the proof of Theorem 15.3.5, multiplying by $(A - \lambda_k I)^{r_k}$ yields

$$\sum_{i=1}^{k-1} (A - \lambda_k I)^{r_k} v_i = \sum_{i=1}^{k-1} v_i' = 0$$

where $v_i' \in V_{\lambda_i}$. Now by induction, each $v_i' = 0$ and so each $v_i = 0$ for $i \leq k-1$. Therefore, the sum, $\sum_{i=1}^{k} v_i$ reduces to v_k and so $v_k = 0$ also.

By Theorem 15.3.5, $\sum_{i=1}^{s} V_{\lambda_i} = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_s} = V$ and by Theorem 15.3.7

$$\{B_1, B_2, \cdots, B_s\}$$

is a basis for V. This proves the corollary.

15.4 Block Diagonal Matrices

In this section the vector space will be $\mathbb{C}^n$ and the linear transformations will be $n \times n$ matrices.

Definition 15.4.1 *Let A and B be two $n \times n$ matrices. Then A is similar to B, written as $A \sim B$ when there exists an invertible matrix, S such that $A = S^{-1}BS$.*

Theorem 15.4.2 *Let A be an $n \times n$ matrix. Letting $\lambda_1, \lambda_2, \cdots, \lambda_r$ be the distinct eigenvalues of A,arranged in any order, there exist square matrices, $P_1, \cdots, P_r$ such that A is similar to the block diagonal matrix,*

$$P = \begin{pmatrix} P_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & P_r \end{pmatrix}$$

in which P_k has the single eigenvalue λ_k. Denoting by r_k the size of P_k it follows that r_k equals the dimension of the generalized eigenspace for λ_k,

$$r_k = \dim\left\{\mathbf{x} : (A-\lambda_k I)^m \mathbf{x} = 0 \text{ for some } m\right\} \equiv \dim\left(V_{\lambda_k}\right)$$

Furthermore, if S is the matrix satisfying $S^{-1}AS = P$, then S is of the form

$$\begin{pmatrix} B_1 & \cdots & B_r \end{pmatrix}$$

where $B_k = \begin{pmatrix} \mathbf{u}_1^k & \cdots & \mathbf{u}_{r_k}^k \end{pmatrix}$ in which the columns, $\left\{\mathbf{u}_1^k, \cdots, \mathbf{u}_{r_k}^k\right\} = D_k$ constitute a basis for V_{λ_k}.

Proof: By Corollary 15.3.8 $\mathbb{C}^n = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_k}$ and a basis for $\mathbb{C}^n$ is $\{D_1, \cdots, D_r\}$ where D_k is a basis for V_{λ_k}.

Let

$$S = \begin{pmatrix} B_1 & \cdots & B_r \end{pmatrix}$$

where the B_i are the matrices described in the statement of the theorem. Then S^{-1} must be of the form

$$S^{-1} = \begin{pmatrix} C_1 \\ \vdots \\ C_r \end{pmatrix}$$

where $C_i B_i = I_{r_i \times r_i}$. Also, if $i \neq j$, then $C_i A B_j = 0$ the last claim holding because $A : V_j \to V_j$ so the columns of AB_j are linear combinations of the columns of B_j and each of these columns is orthogonal to the rows of C_i. Therefore,

$$\begin{aligned} S^{-1}AS &= \begin{pmatrix} C_1 \\ \vdots \\ C_r \end{pmatrix} A \begin{pmatrix} B_1 & \cdots & B_r \end{pmatrix} \\ &= \begin{pmatrix} C_1 \\ \vdots \\ C_r \end{pmatrix} \begin{pmatrix} AB_1 & \cdots & AB_r \end{pmatrix} \\ &= \begin{pmatrix} C_1AB_1 & 0 & \cdots & 0 \\ 0 & C_2AB_2 & \cdots & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & C_rAB_r \end{pmatrix} \end{aligned}$$

and $C_{r_k} A B_{r_k}$ is an $r_k \times r_k$ matrix.

What about the eigenvalues of $C_{r_k} A B_{r_k}$? The only eigenvalue of A restricted to V_{λ_k} is λ_k because if $A\mathbf{x} = \mu\mathbf{x}$ for some $\mathbf{x} \in V_{\lambda_k}$ and $\mu \neq \lambda_k$, then as in Claim 1 of Theorem 15.3.5,

$$(A-\lambda_k I)^{r_k} \mathbf{x} \neq \mathbf{0}$$

contrary to the assumption that $\mathbf{x} \in V_{\lambda_k}$. Suppose then that $C_{r_k} A B_{r_k} \mathbf{x} = \lambda \mathbf{x}$ where $\mathbf{x} \neq \mathbf{0}$. Why is $\lambda = \lambda_k$? Let $\mathbf{y} = B_{r_k}\mathbf{x}$ so $\mathbf{y} \in V_{\lambda_k}$. Then

$$S^{-1}A\mathbf{y} = S^{-1}AS \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{x} \\ \vdots \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \vdots \\ C_{r_k} A B_{r_k} \mathbf{x} \\ \vdots \\ \mathbf{0} \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{x} \\ \vdots \\ \mathbf{0} \end{pmatrix}$$

and so

$$A\mathbf{y} = \lambda S \begin{pmatrix} \mathbf{0} \\ \vdots \\ \mathbf{x} \\ \vdots \\ \mathbf{0} \end{pmatrix} = \lambda \mathbf{y}.$$

Therefore, $\lambda = \lambda_k$ because, as noted above, λ_k is the only eigenvalue of A restricted to V_{λ_k}. Now letting $P_k = C_{r_k} A B_{r_k}$, this proves the theorem.

The above theorem contains a result which is of sufficient importance to state as a corollary.

Corollary 15.4.3 *Let A be an $n \times n$ matrix and let D_k denote a basis for the generalized eigenspace for λ_k. Then $\{D_1, \cdots, D_r\}$ is a basis for $\mathbb{C}^n$.*

More can be said. Recall Theorem 11.2.10 on Page 215. From this theorem, there exist unitary matrices, U_k such that $U_k^* P_k U_k = T_k$ where T_k is an upper triangular matrix of the form

$$\begin{pmatrix} \lambda_k & \cdots & * \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_k \end{pmatrix} \equiv T_k$$

Now let U be the block diagonal matrix defined by

$$U \equiv \begin{pmatrix} U_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & U_r \end{pmatrix}.$$

By Theorem 15.4.2 there exists S such that

$$S^{-1} A S = \begin{pmatrix} P_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & P_r \end{pmatrix}.$$

Therefore,

$$\begin{aligned} U^* S A S U &= \begin{pmatrix} U_1^* & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & U_r^* \end{pmatrix} \begin{pmatrix} P_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & P_r \end{pmatrix} \begin{pmatrix} U_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & U_r \end{pmatrix} \\ &= \begin{pmatrix} U_1^* P_1 U_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & U_r^* P_r U_r \end{pmatrix} = \begin{pmatrix} T_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & T_r \end{pmatrix}. \end{aligned}$$

This proves most of the following corollary of Theorem 15.4.2.

Corollary 15.4.4 *Let A be an $n \times n$ matrix. Then A is similar to an upper triangular, block diagonal matrix of the form*

$$T \equiv \begin{pmatrix} T_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & T_r \end{pmatrix}$$

where T_k is an upper triangular matrix having only λ_k on the main diagonal. The diagonal blocks can be arranged in any order desired. If T_k is an $m_k \times m_k$ matrix, then

$$m_k = \dim\left\{\mathbf{x} : (A - \lambda_k I)^m \mathbf{x} = 0 \text{ for some } m \in \mathbb{N}\right\}.$$

Furthermore, m_k is the multiplicity of λ_k as a zero of the characteristic polynomial of A.

Proof: The only thing which remains is the assertion that m_k equals the multiplicity of λ_k as a zero of the characteristic polynomial. However, this is clear from the observation that since T is similar to A they have the same characteristic polynomial because

$$\begin{aligned}\det(A - \lambda I) &= \det\left(S(T - \lambda I)S^{-1}\right)\\ &= \det(S)\det\left(S^{-1}\right)\det(T - \lambda I)\\ &= \det\left(SS^{-1}\right)\det(T - \lambda I)\\ &= \det(T - \lambda I)\end{aligned}$$

and the observation that since T is upper triangular, the characteristic polynomial of T is of the form

$$\prod_{k=1}^{r} (\lambda_k - \lambda)^{m_k}.$$

The above corollary has tremendous significance especially if it is pushed even further resulting in the Jordan Canonical form. This form involves still more similarity transformations resulting in an especially revealing and simple form for each of the T_k, but the result of the above corollary is sufficient for most applications.

It is significant because it enables one to obtain great understanding of powers of A by using the matrix T. From Corollary 15.4.4 there exists an $n \times n$ matrix, S[2] such that

$$A = S^{-1}TS.$$

Therefore, $A^2 = S^{-1}TSS^{-1}TS = S^{-1}T^2S$ and continuing this way, it follows

$$A^k = S^{-1}T^kS.$$

where T is given in the above corollary. Consider T^k. By block multiplication,

$$T^k = \begin{pmatrix} T_1^k & & 0 \\ & \ddots & \\ 0 & & T_r^k \end{pmatrix}.$$

The matrix, T_s is an $m_s \times m_s$ matrix which is of the form

$$T_s = \begin{pmatrix} \alpha & \cdots & * \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \alpha \end{pmatrix} \tag{15.5}$$

which can be written in the form

$$T_s = D + N$$

for D a multiple of the identity and N an upper triangular matrix with zeros down the main diagonal. Therefore, by the Cayley Hamilton theorem, $N^{m_s} = 0$ because the characteristic

[2]The S here is written as S^{-1} in the corollary.

equation for N is just $\lambda^{m_s} = 0$. Such a transformation is called nilpotent. You can see $N^{m_s} = 0$ directly also, without having to use the Cayley Hamilton theorem. Now since D is just a multiple of the identity, it follows that $DN = ND$. Therefore, the usual binomial theorem may be applied and this yields the following equations for $k \geq m_s$.

$$\begin{aligned} T_s^k &= (D+N)^k = \sum_{j=0}^{k} \binom{k}{j} D^{k-j} N^j \\ &= \sum_{j=0}^{m_s} \binom{k}{j} D^{k-j} N^j, \end{aligned} \tag{15.6}$$

the third equation holding because $N^{m_s} = 0$. Thus T_s^k is of the form

$$T_s^k = \begin{pmatrix} \alpha^k & \cdots & * \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \alpha^k \end{pmatrix}.$$

Lemma 15.4.5 *Suppose T is of the form T_s described above in 15.5 where the constant, α, on the main diagonal is less than one in absolute value. Then*

$$\lim_{k\to\infty} \left(T^k\right)_{ij} = 0.$$

Proof: From 15.6, it follows that for large k, and $j \leq m_s$,

$$\binom{k}{j} \leq \frac{k\,(k-1)\cdots(k-m_s+1)}{m_s!}.$$

Therefore, letting C be the largest value of $\left|\left(N^j\right)_{pq}\right|$ for $0 \leq j \leq m_s$,

$$\left|\left(T^k\right)_{pq}\right| \leq m_s C \left(\frac{k\,(k-1)\cdots(k-m_s+1)}{m_s!}\right) |\alpha|^{k-m_s}$$

which converges to zero as $k \to \infty$. This is most easily seen by applying the ratio test to the series

$$\sum_{k=m_s}^{\infty} \left(\frac{k\,(k-1)\cdots(k-m_s+1)}{m_s!}\right) |\alpha|^{k-m_s}$$

and then noting that if a series converges, then the k^{th} term converges to zero.

15.5 The Matrix Of A Linear Transformation

If V is an n dimensional vector space and $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is a basis for V, there exists a linear map

$$q : \mathbb{F}^n \to V$$

defined as

$$q\,(\mathbf{a}) \equiv \sum_{i=1}^{n} a_i \mathbf{v}_i$$

where

$$\mathbf{a} = \sum_{i=1}^{n} a_i \mathbf{e}_i,$$

for $\mathbf{e}_i$ the standard basis vectors for $\mathbb{F}^n$ consisting of

$$\mathbf{e}_i \equiv \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

where the one is in the i^{th} slot. It is clear that q defined in this way, is one to one, onto, and linear. For $\mathbf{v} \in V$, $q^{-1}(\mathbf{v})$ is a list of scalars called the components of $\mathbf{v}$ with respect to the basis $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$.

Definition 15.5.1 *Given a linear transformation L, mapping V to W, where $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is a basis of V and $\{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ is a basis for W, an $m \times n$ matrix $A = (a_{ij})$is called the matrix of the transformation L with respect to the given choice of bases for V and W, if whenever $\mathbf{v} \in V$, then multiplication of the components of $\mathbf{v}$ by (a_{ij}) yields the components of $L\mathbf{v}$.*

The following diagram is descriptive of the definition. Here q_V and q_W are the maps defined above with reference to the bases, $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $\{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ respectively.

$$\begin{array}{ccccc} & & L & & \\ \{\mathbf{v}_1, \cdots, \mathbf{v}_n\} & V & \rightarrow & W & \{\mathbf{w}_1, \cdots, \mathbf{w}_m\} \\ & q_V \uparrow & \circ & \uparrow q_W & \\ & \mathbb{F}^n & \rightarrow & \mathbb{F}^m & \\ & & A & & \end{array} \tag{15.7}$$

Letting $\mathbf{b} \in \mathbb{F}^n$, this requires

$$\sum_{i,j} a_{ij} b_j \mathbf{w}_i = L \sum_j b_j \mathbf{v}_j = \sum_j b_j L\mathbf{v}_j.$$

Now

$$L\mathbf{v}_j = \sum_i c_{ij} \mathbf{w}_i \tag{15.8}$$

for some choice of scalars c_{ij} because $\{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ is a basis for W. Hence

$$\sum_{i,j} a_{ij} b_j \mathbf{w}_i = \sum_j b_j \sum_i c_{ij} \mathbf{w}_i = \sum_{i,j} c_{ij} b_j \mathbf{w}_i.$$

It follows from the linear independence of $\{\mathbf{w}_1, \cdots, \mathbf{w}_m\}$ that

$$\sum_j a_{ij} b_j = \sum_j c_{ij} b_j$$

for any choice of $\mathbf{b} \in \mathbb{F}^n$ and consequently

$$a_{ij} = c_{ij}$$

where c_{ij} is defined by 15.8. It may help to write 15.8 in the form

$$\begin{pmatrix} L\mathbf{v}_1 & \cdots & L\mathbf{v}_n \end{pmatrix} = \begin{pmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_m \end{pmatrix} C = \begin{pmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_m \end{pmatrix} A \tag{15.9}$$

where $C = (c_{ij}), A = (a_{ij})$.

Example 15.5.2 *Let*

$$V \equiv \{ \text{ polynomials of degree 3 or less} \},$$

$$W \equiv \{ \text{ polynomials of degree 2 or less} \},$$

and $L \equiv D$ where D is the differentiation operator. A basis for V is $\{1, x, x^2, x^3\}$ and a basis for W is $\{1, x, x^2\}$.

What is the matrix of this linear transformation with respect to this basis? Using 15.9,

$$\begin{pmatrix} 0 & 1 & 2x & 3x^2 \end{pmatrix} = \begin{pmatrix} 1 & x & x^2 \end{pmatrix} C.$$

It follows from this that

$$C = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Now consider the important case where $V = \mathbb{F}^n$, $W = \mathbb{F}^m$, and the basis chosen is the standard basis of vectors $\mathbf{e}_i$ described above. Let L be a linear transformation from $\mathbb{F}^n$ to $\mathbb{F}^m$ and let A be the matrix of the transformation with respect to these bases. In this case the coordinate maps q_V and q_W are simply the identity map and the requirement that A is the matrix of the transformation amounts to

$$\pi_i (L\mathbf{b}) = \pi_i (A\mathbf{b})$$

where π_i denotes the map which takes a vector in $\mathbb{F}^m$ and returns the i^{th} entry in the vector, the i^{th} component of the vector with respect to the standard basis vectors. Thus, if the components of the vector in $\mathbb{F}^n$ with respect to the standard basis are $(b_1, \cdots, b_n)$,

$$\mathbf{b} = \begin{pmatrix} b_1 & \cdots & b_n \end{pmatrix}^T = \sum_i b_i \mathbf{e}_i,$$

then

$$\pi_i (L\mathbf{b}) \equiv (L\mathbf{b})_i = \sum_j a_{ij} b_j.$$

What about the situation where different pairs of bases are chosen for V and W? How are the two matrices with respect to these choices related? Consider the following diagram which illustrates the situation.

$$\begin{array}{ccc} \mathbb{F}^n & \xrightarrow{A_2} & \mathbb{F}^m \\ q_2 \downarrow & \circ & p_2 \downarrow \\ V & \xrightarrow{L} & W \\ q_1 \uparrow & \circ & p_1 \uparrow \\ \mathbb{F}^n & \xrightarrow{A_1} & \mathbb{F}^m \end{array}$$

In this diagram q_i and p_i are coordinate maps as described above. From the diagram,

$$p_1^{-1} p_2 A_2 q_2^{-1} q_1 = A_1,$$

where $q_2^{-1} q_1$ and $p_1^{-1} p_2$ are one to one, onto, and linear maps.

Definition 15.5.3 *In the special case where $V = W$ and only one basis is used for $V = W$, this becomes*

$$q_1^{-1} q_2 A_2 q_2^{-1} q_1 = A_1.$$

Letting S be the matrix of the linear transformation $q_2^{-1}q_1$ with respect to the standard basis vectors in $\mathbb{F}^n$,

$$S^{-1}A_2S = A_1. \tag{15.10}$$

When this occurs, A_1 is said to be similar to A_2 and $A \to S^{-1}AS$ is called a similarity transformation.

Here is some terminology.

Definition 15.5.4 *Let S be a set. The symbol, $\sim$ is called an equivalence relation on S if it satisfies the following axioms.*

1. *$x \sim x$ for all $x \in S$. (Reflexive)*
2. *If $x \sim y$ then $y \sim x$. (Symmetric)*
3. *If $x \sim y$ and $y \sim z$, then $x \sim z$. (Transitive)*

Definition 15.5.5 *$[x]$ denotes the set of all elements of S which are equivalent to x and $[x]$ is called the equivalence class determined by x or just the equivalence class of x.*

With the above definition one can prove the following simple theorem which you should do if you have not seen it.

Theorem 15.5.6 *Let $\sim$ be an equivalence class defined on a set, S and let $\mathcal{H}$ denote the set of equivalence classes. Then if $[x]$ and $[y]$ are two of these equivalence classes, either $x \sim y$ and $[x] = [y]$ or it is not true that $x \sim y$ and $[x] \cap [y] = \emptyset$.*

Theorem 15.5.7 *In the vector space of $n \times n$ matrices, define*

$$A \sim B$$

if there exists an invertible matrix S such that

$$A = S^{-1}BS.$$

Then $\sim$ is an equivalence relation and $A \sim B$ if and only if whenever V is an n dimensional vector space, there exists $L \in \mathcal{L}(V,V)$ and bases $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $\{\mathbf{w}_1, \cdots, \mathbf{w}_n\}$ such that A is the matrix of L with respect to $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and B is the matrix of L with respect to $\{\mathbf{w}_1, \cdots, \mathbf{w}_n\}$.

Proof: $A \sim A$ because $S = I$ works in the definition. If $A \sim B$, then $B \sim A$, because

$$A = S^{-1}BS$$

implies

$$B = SAS^{-1}.$$

If $A \sim B$ and $B \sim C$, then

$$A = S^{-1}BS,\ B = T^{-1}CT$$

and so

$$A = S^{-1}T^{-1}CTS = (TS)^{-1}CTS$$

which implies $A \sim C$. This verifies the first part of the conclusion.

Now let V be an n dimensional vector space, $A \sim B$ and pick a basis for V,

$$\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}.$$

Define $L \in \mathcal{L}(V,V)$ by

$$L\mathbf{v}_i \equiv \sum_j a_{ji}\mathbf{v}_j$$

where $A = (a_{ij})$. Then if $B = (b_{ij})$, and $S = (s_{ij})$ is the matrix which provides the similarity transformation,

$$A = S^{-1}BS,$$

between A and B, it follows that

$$L\mathbf{v}_i = \sum_{r,s,j} s_{ir}b_{rs}\left(s^{-1}\right)_{sj}\mathbf{v}_j. \tag{15.11}$$

Now define

$$\mathbf{w}_i \equiv \sum_j \left(s^{-1}\right)_{ij}\mathbf{v}_j.$$

Then from 15.11,

$$\sum_i \left(s^{-1}\right)_{ki} L\mathbf{v}_i = \sum_{i,j,r,s} \left(s^{-1}\right)_{ki} s_{ir}b_{rs}\left(s^{-1}\right)_{sj}\mathbf{v}_j$$

and so

$$L\mathbf{w}_k = \sum_s b_{ks}\mathbf{w}_s.$$

This proves the theorem because the if part of the conclusion was established earlier.

What if the linear transformation consists of multiplication by a matrix A and you want to find the matrix of this linear transformation with respect to another basis? Is there an easy way to do it? The answer is yes.

Proposition 15.5.8 *Let A be an $m \times n$ matrix and let L be the linear transformation which is defined by*

$$L\left(\sum_{k=1}^n x_k\mathbf{e}_k\right) \equiv \sum_{k=1}^n (A\mathbf{e}_k)\, x_k \equiv \sum_{i=1}^m \sum_{k=1}^n A_{ik}x_k\mathbf{e}_i$$

In simple language, to find $L\mathbf{x}$, you multiply on the left of $\mathbf{x}$ by A. Then the matrix M of this linear transformation with respect to the bases $\{\mathbf{u}_1,\cdots,\mathbf{u}_n\}$ for $\mathbb{F}^n$ and $\{\mathbf{w}_1,\cdots,\mathbf{w}_m\}$ for $\mathbb{F}^m$ is given by

$$M = \begin{pmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_m \end{pmatrix}^{-1} A \begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{pmatrix}$$

where $\begin{pmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_m \end{pmatrix}$ is the $m \times m$ matrix which has $\mathbf{w}_j$ as its j^{th} column.

Proof: Consider the following diagram.

$$\begin{array}{ccccc} & & L & & \\ \{\mathbf{u}_1,\cdots,\mathbf{u}_n\} & \mathbb{F}^n & \rightarrow & \mathbb{F}^m & \{\mathbf{w}_1,\cdots,\mathbf{w}_m\} \\ & q_V \uparrow & \circ & \uparrow q_W & \\ & \mathbb{F}^n & \rightarrow & \mathbb{F}^m & \\ & & M & & \end{array}$$

Here the coordinate maps are defined in the usual way. Thus

$$q_V\begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix}^T \equiv \sum_{i=1}^n x_i\mathbf{u}_i.$$

Therefore, q_V can be considered the same as multiplication of a vector in $\mathbb{F}^n$ on the left by the matrix

$$\begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{pmatrix}.$$

Similar considerations apply to q_W. Thus it is desired to have the following for an arbitrary $\mathbf{x} \in \mathbb{F}^n$.

$$A \begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{pmatrix} \mathbf{x} = \begin{pmatrix} \mathbf{w}_1 & \cdots & \mathbf{w}_n \end{pmatrix} M\mathbf{x}$$

Therefore, the conclusion of the proposition follows. This proves the proposition.

Definition 15.5.9 *An $n \times n$ matrix, A, is diagonalizable if there exists an invertible $n \times n$ matrix, S such that $S^{-1}AS = D$, where D is a diagonal matrix. Thus D has zero entries everywhere except on the main diagonal. Write* $\operatorname{diag}(\lambda_1 \cdots, \lambda_n)$ *to denote the diagonal matrix having the λ_i down the main diagonal.*

Which matrices are diagonalizable?

Theorem 15.5.10 *Let A be an $n \times n$ matrix. Then A is diagonalizable if and only if $\mathbb{F}^n$ has a basis of eigenvectors of A. In this case, S of Definition 15.5.9 consists of the $n \times n$ matrix whose columns are the eigenvectors of A and $D = \operatorname{diag}(\lambda_1, \cdots, \lambda_n)$.*

Proof: Suppose first that $\mathbb{F}^n$ has a basis of eigenvectors, $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ where $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$. Then let S denote the matrix $(\mathbf{v}_1 \cdots \mathbf{v}_n)$ and let $S^{-1} \equiv \begin{pmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{pmatrix}$ where $\mathbf{u}_i^T \mathbf{v}_j = \delta_{ij} \equiv \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$. S^{-1} exists because S has rank n. Then from block multiplication,

$$S^{-1}AS = \begin{pmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{pmatrix} (A\mathbf{v}_1 \cdots A\mathbf{v}_n)$$

$$= \begin{pmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{pmatrix} (\lambda_1 \mathbf{v}_1 \cdots \lambda_n \mathbf{v}_n)$$

$$= \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix} = D.$$

Next suppose A is diagonalizable so $S^{-1}AS = D \equiv \operatorname{diag}(\lambda_1, \cdots, \lambda_n)$. Then the columns of S form a basis because S^{-1} is given to exist. It only remains to verify that these columns of A are eigenvectors. But letting $S = (\mathbf{v}_1 \cdots \mathbf{v}_n)$, $AS = SD$ and so $(A\mathbf{v}_1 \cdots A\mathbf{v}_n) = (\lambda_1 \mathbf{v}_1 \cdots \lambda_n \mathbf{v}_n)$ which shows that $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$. This proves the theorem.

It makes sense to speak of the determinant of a linear transformation as described in the following corollary.

Corollary 15.5.11 *Let $L \in \mathcal{L}(V, V)$ where V is an n dimensional vector space and let A be the matrix of this linear transformation with respect to a basis on V. Then it is possible to define*

$$\det(L) \equiv \det(A).$$

Proof: Each choice of basis for V determines a matrix for L with respect to the basis. If A and B are two such matrices, it follows from Theorem 15.5.7 that

$$A = S^{-1}BS$$

and so

$$\det(A) = \det\left(S^{-1}\right)\det(B)\det(S).$$

But

$$1 = \det(I) = \det\left(S^{-1}S\right) = \det(S)\det\left(S^{-1}\right)$$

and so

$$\det(A) = \det(B)$$

which proves the corollary.

Definition 15.5.12 *Let $A \in \mathcal{L}(X,Y)$ where X and Y are finite dimensional vector spaces. Define* $\operatorname{rank}(A)$ *to equal the dimension of $A(X)$.*

The following theorem explains how the rank of A is related to the rank of the matrix of A.

Theorem 15.5.13 *Let $A \in \mathcal{L}(X,Y)$. Then* $\operatorname{rank}(A) = \operatorname{rank}(M)$ *where M is the matrix of A taken with respect to a pair of bases for the vector spaces X, and Y.*

Proof: Recall the diagram which describes what is meant by the matrix of A. Here the two bases are as indicated.

$$\begin{array}{ccccc} \{v_1,\cdots,v_n\} & X & \underrightarrow{A} & Y & \{w_1,\cdots,w_m\} \\ & q_X\uparrow & \circ & \uparrow q_Y & \\ & \mathbb{F}^n & \underrightarrow{M} & \mathbb{F}^m & \end{array}$$

Let $\{z_1,\cdots,z_r\}$ be a basis for $A(X)$. Then since the linear transformation, q_Y is one to one and onto, $\left\{q_Y^{-1}z_1,\cdots,q_Y^{-1}z_r\right\}$ is a linearly independent set of vectors in $\mathbb{F}^m$. Let $Au_i = z_i$. Then

$$Mq_X^{-1}u_i = q_Y^{-1}z_i$$

and so the dimension of $M(\mathbb{F}^n) \geq r$. Now if $M(\mathbb{F}^n) < r$ then there exists

$$\mathbf{y} \in M(\mathbb{F}^n) \setminus \operatorname{span}\left\{q_Y^{-1}z_1,\cdots,q_Y^{-1}z_r\right\}.$$

But then there exists $\mathbf{x} \in \mathbb{F}^n$ with $M\mathbf{x} = \mathbf{y}$. Hence

$$\mathbf{y} = M\mathbf{x} = q_Y^{-1}Aq_X\mathbf{x} \in \operatorname{span}\left\{q_Y^{-1}z_1,\cdots,q_Y^{-1}z_r\right\}$$

a contradiction. This proves the theorem.

The following result is a summary of many concepts.

Theorem 15.5.14 *Let $L \in \mathcal{L}(V,V)$ where V is a finite dimensional vector space. Then the following are equivalent.*

1. *L is one to one.*
2. *L maps a basis to a basis.*
3. *L is onto.*

4. $\det(L) \neq 0$

5. *If* $Lv = 0$ *then* $v = 0$.

Proof: Suppose first L is one to one and let $\{v_i\}_{i=1}^n$ be a basis. Then if $\sum_{i=1}^n c_i L v_i = 0$ it follows $L\left(\sum_{i=1}^n c_i v_i\right) = 0$ which means that since $L(0) = 0$, and L is one to one, it must be the case that $\sum_{i=1}^n c_i v_i = 0$. Since $\{v_i\}$ is a basis, each $c_i = 0$ which shows $\{Lv_i\}$ is a linearly independent set. Since there are n of these, it must be that this is a basis.

Now suppose 2.). Then letting $\{v_i\}$ be a basis, and $y \in V$, it follows from part 2.) that there are constants, $\{c_i\}$ such that $y = \sum_{i=1}^n c_i L v_i = L\left(\sum_{i=1}^n c_i v_i\right)$. Thus L is onto. It has been shown that 2.) implies 3.).

Now suppose 3.). Then the operation consisting of multiplication by the matrix of L, M_L, must be onto. However, the vectors in $\mathbb{F}^n$ so obtained, consist of linear combinations of the columns of M_L. Therefore, the column rank of M_L is n. By Theorem 7.5.7 this equals the determinant rank and so $\det(M_L) \equiv \det(L) \neq 0$.

Now assume 4.) If $Lv = 0$ for some $v \neq 0$, it follows that $M_L \mathbf{x} = 0$ for some $\mathbf{x} \neq \mathbf{0}$. Therefore, the columns of M_L are linearly dependent and so by Theorem 7.5.7, $\det(M_L) = \det(L) = 0$ contrary to 4.). Therefore, 4.) implies 5.).

Now suppose 5.) and suppose $Lv = Lw$. Then $L(v - w) = 0$ and so by 5.), $v - w = 0$ showing that L is one to one. This proves the theorem.

Also it is important to note that composition of linear transformation corresponds to multiplication of the matrices. Consider the following diagram.

$$\begin{array}{ccccc} X & \underrightarrow{A} & Y & \underrightarrow{B} & Z \\ q_X \uparrow & \circ & \uparrow q_Y & \circ & \uparrow q_Z \\ \mathbb{F}^n & \underrightarrow{M_A} & \mathbb{F}^m & \underrightarrow{M_B} & \mathbb{F}^p \end{array}$$

where A and B are two linear transformations, $A \in \mathcal{L}(X, Y)$ and $B \in \mathcal{L}(Y, Z)$. Then $B \circ A \in \mathcal{L}(X, Z)$ and so it has a matrix with respect to bases given on X and Z, the coordinate maps for these bases being q_X and q_Z respectively. Then

$$B \circ A = q_Z M_B q_Y q_Y^{-1} M_A q_X^{-1} = q_Z M_B M_A q_X^{-1}.$$

But this shows that $M_B M_A$ plays the role of $M_{B \circ A}$, the matrix of $B \circ A$. Hence the matrix of $B \circ A$ equals the product of the two matrices M_A and M_B. Of course it is interesting to note that although $M_{B \circ A}$ must be unique, the matrices, M_B and M_A are not unique, depending on the basis chosen for Y.

Theorem 15.5.15 *The matrix of the composition of linear transformations equals the product of the matrices of these linear transformations.*

15.5.1 Some Geometrically Defined Linear Transformations

If T is any linear transformation which maps $\mathbb{F}^n$ to $\mathbb{F}^m$, there is always an $m \times n$ matrix, A with the property that

$$A\mathbf{x} = T\mathbf{x} \tag{15.12}$$

for all $\mathbf{x} \in \mathbb{F}^n$. You simply take the matrix of the linear transformation with respect to the standard basis. What is the form of A? Suppose $T : \mathbb{F}^n \to \mathbb{F}^m$ is a linear transformation and you want to find the matrix defined by this linear transformation as described in 15.12. Then if $\mathbf{x} \in \mathbb{F}^n$ it follows

$$\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i$$

where $\mathbf{e}_i$ is the vector which has zeros in every slot but the i^{th} and a 1 in this slot. Then since T is linear,

$$\begin{aligned} T\mathbf{x} &= \sum_{i=1}^{n} x_i T(\mathbf{e}_i) \\ &= \begin{pmatrix} | & & | \\ T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \\ | & & | \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\ &\equiv A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \end{aligned}$$

and so you see that the matrix desired is obtained from letting the i^{th} column equal $T(\mathbf{e}_i)$. This proves the following theorem.

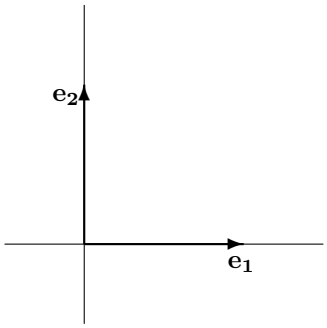
Theorem 15.5.16 *Let T be a linear transformation from $\mathbb{F}^n$ to $\mathbb{F}^m$. Then the matrix, A satisfying 15.12 is given by*

$$\begin{pmatrix} | & & | \\ T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \\ | & & | \end{pmatrix}$$

where $T\mathbf{e}_i$ is the i^{th} column of A.

Example 15.5.17 *Determine the matrix for the transformation mapping $\mathbb{R}^2$ to $\mathbb{R}^2$ which consists of rotating every vector counter clockwise through an angle of θ.*

Let $\mathbf{e}_1 \equiv \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\mathbf{e}_2 \equiv \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. These identify the geometric vectors which point along the positive x axis and positive y axis as shown.



From Theorem 15.5.16, you only need to find $T\mathbf{e}_1$ and $T\mathbf{e}_2$, the first being the first column of the desired matrix, A and the second being the second column. From drawing a picture and doing a little geometry, you see that

$$T\mathbf{e}_1 = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, T\mathbf{e}_2 = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}.$$

Therefore, from Theorem 15.5.16,

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Example 15.5.18 *Find the matrix of the linear transformation which is obtained by first rotating all vectors through an angle of ϕ and then through an angle θ. Thus you want the linear transformation which rotates all angles through an angle of $\theta+\phi$.*

Let $T_{\theta+\phi}$ denote the linear transformation which rotates every vector through an angle of $\theta+\phi$. Then to get $T_{\theta+\phi}$, you could first do T_ϕ and then do T_θ where T_ϕ is the linear transformation which rotates through an angle of ϕ and T_θ is the linear transformation which rotates through an angle of θ. Denoting the corresponding matrices by $A_{\theta+\phi}$, A_ϕ, and A_θ, you must have for every $\mathbf{x}$

$$A_{\theta+\phi}\mathbf{x} = T_{\theta+\phi}\mathbf{x} = T_\theta T_\phi \mathbf{x} = A_\theta A_\phi \mathbf{x}.$$

Consequently, you must have

$$\begin{aligned} A_{\theta+\phi} &= \begin{pmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{pmatrix} = A_\theta A_\phi \\ &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}. \end{aligned}$$

Therefore,

$$\begin{pmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{pmatrix} = \begin{pmatrix} \cos\theta\cos\phi - \sin\theta\sin\phi & -\cos\theta\sin\phi - \sin\theta\cos\phi \\ \sin\theta\cos\phi + \cos\theta\sin\phi & \cos\theta\cos\phi - \sin\theta\sin\phi \end{pmatrix}.$$

Don't these look familiar? They are the usual trig. identities for the sum of two angles derived here using linear algebra concepts.

Example 15.5.19 *Find the matrix of the linear transformation which rotates vectors in $\mathbb{R}^3$ counterclockwise about the positive z axis.*

Let T be the name of this linear transformation. In this case, $T\mathbf{e}_3 = \mathbf{e}_3, T\mathbf{e}_1 = (\cos\theta, \sin\theta, 0)^T$, and $T\mathbf{e}_2 = (-\sin\theta, \cos\theta, 0)^T$. Therefore, the matrix of this transformation is just

$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{15.13}$$

In Physics it is important to consider the work done by a force field on an object. This involves the concept of projection onto a vector. Suppose you want to find the projection of a vector, $\mathbf{v}$ onto the given vector, $\mathbf{u}$, denoted by $\mathrm{proj}_{\mathbf{u}}(\mathbf{v})$ This is done using the dot product as follows.

$$\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) = \left(\frac{\mathbf{v}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u}$$

Because of properties of the dot product, the map $\mathbf{v} \to \mathrm{proj}_{\mathbf{u}}(\mathbf{v})$ is linear,

$$\begin{aligned} \mathrm{proj}_{\mathbf{u}}(\alpha\mathbf{v}+\beta\mathbf{w}) &= \left(\frac{\alpha\mathbf{v}+\beta\mathbf{w}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u} = \alpha\left(\frac{\mathbf{v}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u} + \beta\left(\frac{\mathbf{w}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\right)\mathbf{u} \\ &= \alpha\,\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) + \beta\,\mathrm{proj}_{\mathbf{u}}(\mathbf{w}). \end{aligned}$$

Example 15.5.20 *Let the projection map be defined above and let $\mathbf{u} = (1,2,3)^T$. Find the matrix of this linear transformation with respect to the usual basis.*

You can find this matrix in the same way as in earlier examples. $\text{proj}_{\mathbf{u}}(\mathbf{e}_i)$ gives the i^{th} column of the desired matrix. Therefore, it is only necessary to find

$$\text{proj}_{\mathbf{u}}(\mathbf{e}_i) \equiv \left(\frac{\mathbf{e}_i \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\right) \mathbf{u}$$

For the given vector in the example, this implies the columns of the desired matrix are

$$\frac{1}{14}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \frac{2}{14}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \frac{3}{14}\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Hence the matrix is

$$\frac{1}{14}\begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix}.$$

Example 15.5.21 *Find the matrix of the linear transformation which reflects all vectors in $\mathbb{R}^3$ through the xz plane.*

As illustrated above, you just need to find $T\mathbf{e}_i$ where T is the name of the transformation. But $T\mathbf{e}_1 = \mathbf{e}_1, T\mathbf{e}_3 = \mathbf{e}_3$, and $T\mathbf{e}_2 = -\mathbf{e}_2$ so the matrix is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Example 15.5.22 *Find the matrix of the linear transformation which first rotates counter clockwise about the positive z axis and then reflects through the xz plane.*

This linear transformation is just the composition of two linear transformations having matrices

$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

respectively. Thus the matrix desired is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ -\sin\theta & -\cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

15.5.2 Rotations About A Given Vector

As an application, I will consider the problem of rotating counter clockwise about a given unit vector which is possibly not one of the unit vectors in coordinate directions. First consider a pair of perpendicular unit vectors, $\mathbf{u}_1$ and $\mathbf{u}_2$ and the problem of rotating in the counterclockwise direction about $\mathbf{u}_3$ where $\mathbf{u}_3 = \mathbf{u}_1 \times \mathbf{u}_2$ so that $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ forms a right handed orthogonal coordinate system. Let T denote the desired rotation. Then

$$T(a\mathbf{u}_1 + b\mathbf{u}_2 + c\mathbf{u}_3) = aT\mathbf{u}_1 + bT\mathbf{u}_2 + cT\mathbf{u}_3$$

$$= (a\cos\theta - b\sin\theta)\,\mathbf{u}_1 + (a\sin\theta + b\cos\theta)\,\mathbf{u}_2 + c\mathbf{u}_3.$$

Thus in terms of the basis $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$, the matrix of this transformation is

$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

I want to write this transformation in terms of the usual basis vectors, $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$. From Proposition 15.5.8, if A is this matrix,

$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{pmatrix}^{-1} A \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{pmatrix}$$

and so you can solve for A if you know the $\mathbf{u}_i$.

Suppose the unit vector about which the counterclockwise rotation takes place is (a, b, c). Then I obtain vectors, $\mathbf{u}_1$ and $\mathbf{u}_2$ such that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is a right handed orthogonal system with $\mathbf{u}_3 = (a, b, c)$ and then use the above result. It is of course somewhat arbitrary how this is accomplished. I will assume, however that $|c| \neq 1$ since otherwise you are looking at either clockwise or counter clockwise rotation about the positive z axis and this is a problem which has been dealt with earlier. (If $c = -1$, it amounts to clockwise rotation about the positive z axis while if $c = 1$, it is counterclockwise rotation about the positive z axis.) Then let $\mathbf{u}_3 = (a, b, c)$ and $\mathbf{u}_2 \equiv \frac{1}{\sqrt{a^2+b^2}}(b, -a, 0)$. This one is perpendicular to $\mathbf{u}_3$. If $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is to be a right hand system it is necessary to have

$$\mathbf{u}_1 = \mathbf{u}_2 \times \mathbf{u}_3 = \frac{1}{\sqrt{(a^2+b^2)\,(a^2+b^2+c^2)}}\left(-ac, -bc, a^2+b^2\right)$$

Now recall that $\mathbf{u}_3$ is a unit vector and so the above equals

$$\frac{1}{\sqrt{(a^2+b^2)}}\left(-ac, -bc, a^2+b^2\right)$$

Then from the above, A is given by

$$\begin{pmatrix} \frac{-ac}{\sqrt{(a^2+b^2)}} & \frac{b}{\sqrt{a^2+b^2}} & a \\ \frac{-bc}{\sqrt{(a^2+b^2)}} & \frac{-a}{\sqrt{a^2+b^2}} & b \\ \sqrt{a^2+b^2} & 0 & c \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot$$

$$\begin{pmatrix} \frac{-ac}{\sqrt{(a^2+b^2)}} & \frac{b}{\sqrt{a^2+b^2}} & a \\ \frac{-bc}{\sqrt{(a^2+b^2)}} & \frac{-a}{\sqrt{a^2+b^2}} & b \\ \sqrt{a^2+b^2} & 0 & c \end{pmatrix}^{-1}$$

Of course the matrix is an orthogonal matrix so it is easy to take the inverse by simply taking the transpose. Then doing the computation and then some simplification yields

$$= \begin{pmatrix} a^2 + \left(1-a^2\right)\cos\theta & ab\,(1-\cos\theta) - c\sin\theta & ac\,(1-\cos\theta) + b\sin\theta \\ ab\,(1-\cos\theta) + c\sin\theta & b^2 + \left(1-b^2\right)\cos\theta & bc\,(1-\cos\theta) - a\sin\theta \\ ac\,(1-\cos\theta) - b\sin\theta & bc\,(1-\cos\theta) + a\sin\theta & c^2 + \left(1-c^2\right)\cos\theta \end{pmatrix}. \quad (15.14)$$

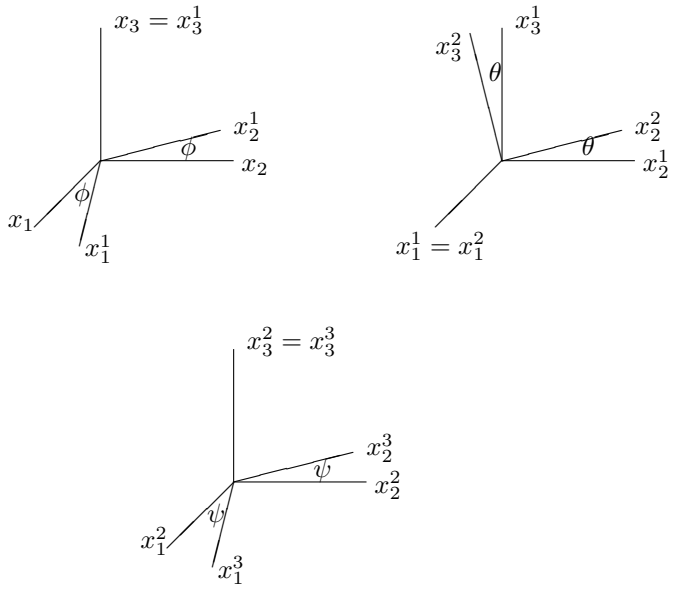
With this, it is clear how to rotate clockwise about the the unit vector, (a,b,c). Just rotate counter clockwise through an angle of $-\theta$. Thus the matrix for this clockwise roation is just

$$= \begin{pmatrix} a^2 + \left(1-a^2\right)\cos\theta & ab\left(1-\cos\theta\right) + c\sin\theta & ac\left(1-\cos\theta\right) - b\sin\theta \\ ab\left(1-\cos\theta\right) - c\sin\theta & b^2 + \left(1-b^2\right)\cos\theta & bc\left(1-\cos\theta\right) + a\sin\theta \\ ac\left(1-\cos\theta\right) + b\sin\theta & bc\left(1-\cos\theta\right) - a\sin\theta & c^2 + \left(1-c^2\right)\cos\theta \end{pmatrix}.$$

In deriving 15.14 it was assumed that $c \neq \pm 1$ but even in this case, it gives the correct answer. Suppose for example that $c = 1$ so you are rotating in the counter clockwise direction about the positive z axis. Then a, b are both equal to zero and 15.14 reduces to 15.13.

15.5.3 The Euler Angles

An important application of the above theory is to the Euler angles, important in the mechanics of rotating bodies. Lagrange studied these things back in the 1700's. To describe the Euler angles consider the following picture in which x_1, x_2 and x_3 are the usual coordinate axes fixed in space and the axes labeled with a superscript denote other coordinate axes. Here is the picture.



We obtain ϕ by rotating counter clockwise about the fixed x_3 axis. Thus this rotation has the matrix

$$\begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \equiv M_1(\phi)$$

Next rotate counter clockwise about the x_1^1 axis which results from the first rotation through an angle of θ. Thus it is desired to rotate counter clockwise through an angle θ about the

unit vector

$$\begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos\phi \\ \sin\phi \\ 0 \end{pmatrix}.$$

Therefore, in 15.14, $a = \cos\phi, b = \sin\phi$, and $c = 0$. It follows the matrix of this transformation with respect to the usual basis is

$$\begin{pmatrix} \cos^2\phi + \sin^2\phi\cos\theta & \cos\phi\sin\phi\left(1-\cos\theta\right) & \sin\phi\sin\theta \\ \cos\phi\sin\phi\left(1-\cos\theta\right) & \sin^2\phi + \cos^2\phi\cos\theta & -\cos\phi\sin\theta \\ -\sin\phi\sin\theta & \cos\phi\sin\theta & \cos\theta \end{pmatrix} \equiv M_2\left(\phi,\theta\right)$$

Finally, we rotate counter clockwise about the positive x_3^2 axis by ψ. The vector in the positive x_3^1 axis is the same as the vector in the fixed x_3 axis. Thus the unit vector in the positive direction of the x_3^2 axis is

$$\begin{pmatrix} \cos^2\phi + \sin^2\phi\cos\theta & \cos\phi\sin\phi\left(1-\cos\theta\right) & \sin\phi\sin\theta \\ \cos\phi\sin\phi\left(1-\cos\theta\right) & \sin^2\phi + \cos^2\phi\cos\theta & -\cos\phi\sin\theta \\ -\sin\phi\sin\theta & \cos\phi\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$
$$= \begin{pmatrix} \cos^2\phi + \sin^2\phi\cos\theta \\ \cos\phi\sin\phi\left(1-\cos\theta\right) \\ -\sin\phi\sin\theta \end{pmatrix} = \begin{pmatrix} \cos^2\phi + \sin^2\phi\cos\theta \\ \cos\phi\sin\phi\left(1-\cos\theta\right) \\ -\sin\phi\sin\theta \end{pmatrix}$$

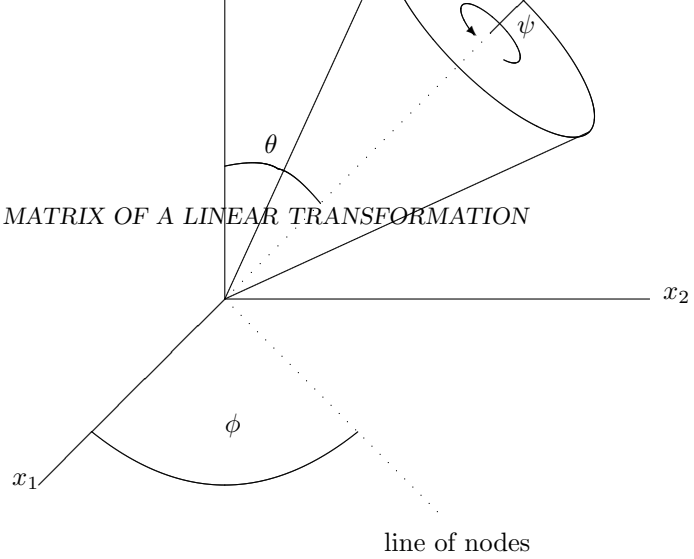
and it is desired to rotate counter clockwise through an angle of ψ about this vector. Thus, in this case,

$$a = \cos^2\phi + \sin^2\phi\cos\theta, b = \cos\phi\sin\phi\left(1-\cos\theta\right), c = -\sin\phi\sin\theta.$$

and you could substitute in to the formula of Theorem 15.14 and obtain a matrix which represents the linear transformation obtained by rotating counter clockwise about the positive x_3^2 axis, $M_3\left(\phi,\theta,\psi\right)$. Then what would be the matrix with respect to the usual basis for the linear transformation which is obtained as a composition of the three just described? By Theorem 15.5.15, this matrix equals the product of these three,

$$M_3\left(\phi,\theta,\psi\right) M_2\left(\phi,\theta\right) M_1\left(\phi\right).$$

I leave the details to you. There are procedures due to Lagrange which will allow you to write differential equations for the Euler angles in a rotating body. To give an idea how these angles apply, consider the following picture.



This is as far as I will go on this topic. The point is, it is possible to give a systematic description in terms of matrix multiplication of a very elaborate geometrical description of a composition of linear transformations. You see from the picture it is possible to describe the motion of the spinning top shown in terms of these Euler angles. I think you can also see that the end result would be pretty horrendous but this is because it involves using the basis corresponding to a fixed in space coordinate system. You wouldn't do this for the application to a spinning top.

Not surprisingly, this also has applications to computer graphics.

The Jordan Canonical Form*

Recall Corollary 15.4.4. For convenience, this corollary is stated below.

Corollary A.0.23 *Let A be an $n \times n$ matrix. Then A is similar to an upper triangular, block diagonal matrix of the form*

$$T \equiv \begin{pmatrix} T_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & T_r \end{pmatrix}$$

where T_k is an upper triangular matrix having only λ_k on the main diagonal. The diagonal blocks can be arranged in any order desired. If T_k is an $m_k \times m_k$ matrix, then

$$m_k = \dim \left\{ \mathbf{x} : (A - \lambda_k I)^m \mathbf{x} = 0 \text{ for some } m \in \mathbb{N} \right\}.$$

The Jordan Canonical form involves a further reduction in which the upper triangular matrices, T_k assume a particularly revealing and simple form.

Definition A.0.24 *$J_k(\alpha)$ is a Jordan block if it is a $k \times k$ matrix of the form*

$$J_k(\alpha) = \begin{pmatrix} \alpha & 1 & & 0 \\ 0 & \ddots & \ddots & \\ \vdots & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & \alpha \end{pmatrix}$$

In words, there is an unbroken string of ones down the super diagonal and the number, α filling every space on the main diagonal with zeros everywhere else. A matrix is strictly upper triangular if it is of the form

$$\begin{pmatrix} 0 & * & * \\ \vdots & \ddots & * \\ 0 & \cdots & 0 \end{pmatrix},$$

where there are zeroes on the main diagonal and below the main diagonal.

The Jordan canonical form involves each of the upper triangular matrices in the conclusion of Corollary 15.4.4 being a block diagonal matrix with the blocks being Jordan blocks in which the size of the blocks decreases from the upper left to the lower right. The idea is to show that every square matrix is similar to a unique such matrix which is in Jordan canonical form.

Note that in the conclusion of Corollary 15.4.4 each of the triangular matrices is of the form $\alpha I + N$ where N is a strictly upper triangular matrix. The existence of the Jordan canonical form follows quickly from the following lemma.

Lemma A.0.25 *Let N be an $n \times n$ matrix which is strictly upper triangular. Then there exists an invertible matrix, S such that*

$$S^{-1}NS = \begin{pmatrix} J_{r_1}(0) & & & 0 \\ & J_{r_2}(0) & & \\ & & \ddots & \\ 0 & & & J_{r_s}(0) \end{pmatrix}$$

where $r_1 \geq r_2 \geq \cdots \geq r_s \geq 1$ and $\sum_{i=1}^{s} r_i = n$.

Proof: First note the only eigenvalue of N is 0. Let $\mathbf{v}_1$ be an eigenvector. Then $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r\}$ is called a chain if $N\mathbf{v}_{k+1} = \mathbf{v}_k$ for all $k = 1, 2, \cdots, r$ and $\mathbf{v}_1$ is an eigenvector. It will be called a maximal chain if there is no solution, $\mathbf{v}$, to the equation, $N\mathbf{v} = \mathbf{v}_r$.

Claim 1: The vectors in any chain are linearly independent and for $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r\}$ a chain based on $\mathbf{v}_1$,

$$N : \operatorname{span}(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r) \to \operatorname{span}(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r). \tag{1.1}$$

Also if $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r\}$ is a chain, then $r \leq n$.

Proof: First note that 1.1 is obvious because

$$N \sum_{i=1}^{r} c_i \mathbf{v}_i = \sum_{i=2}^{r} c_i \mathbf{v}_{i-1}.$$

It only remains to verify the vectors of a chain are independent. Suppose then

$$\sum_{k=1}^{r} c_k \mathbf{v}_k = 0.$$

Do N^{r-1} to it to conclude $c_r = 0$. Next do N^{r-2} to it to conclude $c_{r-1} = 0$ and continue this way. Now it is obvious $r \leq n$ because the chain is independent. This proves the claim.

Consider the set of all chains based on eigenvectors. Since all have total length no larger than n it follows there exists one which has maximal length, $\left\{\mathbf{v}_1^1, \cdots, \mathbf{v}_{r_1}^1\right\} \equiv B_1$. If $\operatorname{span}(B_1)$ contains all eigenvectors of N, then stop. Otherwise, consider all chains based on eigenvectors not in $\operatorname{span}(B_1)$ and pick one, $B_2 \equiv \left\{\mathbf{v}_1^2, \cdots, \mathbf{v}_{r_2}^2\right\}$ which is as long as possible. Thus $r_2 \leq r_1$. If $\operatorname{span}(B_1, B_2)$ contains all eigenvectors of N, stop. Otherwise, consider all chains based on eigenvectors not in $\operatorname{span}(B_1, B_2)$ and pick one, $B_3 \equiv \left\{\mathbf{v}_1^3, \cdots, \mathbf{v}_{r_3}^3\right\}$ such that r_3 is as large as possible. Continue this way. Thus $r_k \geq r_{k+1}$.

Claim 2: The above process terminates with a finite list of chains, $\{B_1, \cdots, B_s\}$ because for any k, $\{B_1, \cdots, B_k\}$ is linearly independent.

Proof of Claim 2: The claim is true if $k = 1$. This follows from Claim 1. Suppose it is true for $k-1, k \geq 2$. Then $\{B_1, \cdots, B_{k-1}\}$ is linearly independent. Suppose

$$\sum_{q=1}^{p} c_q \mathbf{w}_q = \mathbf{0}, \, c_q \neq 0$$

where the $\mathbf{w}_q$ come from $\{B_1, \cdots, B_{k-1}, B_k\}$. By induction, some of these $\mathbf{w}_q$ must come from B_k. Let $\mathbf{v}_i^k$ be the one for which i is as large as possible. Then do N^{i-1} to both sides to obtain $\mathbf{v}_1^k$, the eigenvector upon which the chain B_k is based, is a linear combination of $\{B_1, \cdots, B_{k-1}\}$ contrary to the construction. Since $\{B_1, \cdots, B_k\}$ is linearly independent, the process terminates. This proves the claim.

Claim 3: Suppose $N\mathbf{w} = \mathbf{0}$. ($\mathbf{w}$ is an eigenvector) Then there exists scalars, c_i such that

$$\mathbf{w} = \sum_{i=1}^{s} c_i \mathbf{v}_1^i.$$

Recall that $\mathbf{v}_1^i$ is the eigenvector in the i^{th} chain on which this chain is based.

Proof of Claim 3: From the construction, $\mathbf{w} \in \text{span}\,(B_1, \cdots, B_s)$ since otherwise, it could serve as a base for another chain. Therefore,

$$\mathbf{w} = \sum_{i=1}^{s} \sum_{k=1}^{r_i} c_i^k \mathbf{v}_k^i.$$

Now apply N to both sides.

$$\mathbf{0} = \sum_{i=1}^{s} \sum_{k=2}^{r_i} c_i^k \mathbf{v}_{k-1}^i$$

and so by **Claim 2,** $c_i^k = 0$ if $k \geq 2$. Therefore,

$$\mathbf{w} = \sum_{i=1}^{s} c_i^1 \mathbf{v}_1^i$$

and this proves the claim.

It remains to verify that $\text{span}\,(B_1, \cdots, B_s) = \mathbb{F}^n$. Suppose $\mathbf{w} \notin \text{span}\,(B_1, \cdots, B_s)$. By **Claim 3** this implies $\mathbf{w}$ is not an eigenvector since all the eigenvectors are in $\text{span}\,(B_1, \cdots, B_s)$. Since $N^n = 0$, there exists a smallest integer, $k \geq 2$ such that $N^k \mathbf{w} = \mathbf{0}$ but $N^{k-1}\mathbf{w} \neq \mathbf{0}$. Then $k \leq \min(r_1, \cdots, r_s)$ because there exists a chain of length k based on the eigenvector, $N^{k-1}\mathbf{w}$, namely

$$N^{k-1}\mathbf{w}, N^{k-2}\mathbf{w}, N^{k-3}\mathbf{w}, \cdots, \mathbf{w}$$

and this chain must be no longer than the preceding chains because of the construction in which a longest possible chain was chosen at each step. Since $N^{k-1}\mathbf{w}$ is an eigenvector, it follows from **Claim 3** that

$$N^{k-1}\mathbf{w} = \sum_{i=1}^{s} c_i \mathbf{v}_1^i = \sum_{i=1}^{s} c_i N^{k-1} \mathbf{v}_k^i.$$

Therefore,

$$N^{k-1}\left(\mathbf{w} - \sum_{i=1}^{s} c_i \mathbf{v}_k^i\right) = \mathbf{0}$$

and so,

$$NN^{k-2}\left(\mathbf{w}-\sum_{i=1}^{s}c_i\mathbf{v}_k^i\right)=\mathbf{0}$$

which implies $N^{k-2}\left(\mathbf{w}-\sum_{i=1}^{s}c_i\mathbf{v}_k^i\right)$ is an eigenvector and so by **Claim 3** there exist d_i such that

$$N^{k-2}\left(\mathbf{w}-\sum_{i=1}^{s}c_i\mathbf{v}_k^i\right)=\sum_{i=1}^{s}d_i\mathbf{v}_1^i=\sum_{i=1}^{s}d_iN^{k-2}\mathbf{v}_{k-1}^i$$

and so

$$N^{k-2}\left(\mathbf{w}-\sum_{i=1}^{s}c_i\mathbf{v}_k^i-\sum_{i=1}^{s}d_i\mathbf{v}_{k-1}^i\right)=\mathbf{0}.$$

Continuing this way it follows that for each $j<k$, there exists a vector, $\mathbf{z}_j\in\operatorname{span}(B_1,\cdots,B_s)$ such that

$$N^{k-j}(\mathbf{w}-\mathbf{z}_j)=\mathbf{0}.$$

In particular, taking $j=(k-1)$ yields

$$N(\mathbf{w}-\mathbf{z}_{k-1})=\mathbf{0}$$

and now using **Claim 3** again yields $\mathbf{w}\in\operatorname{span}(B_1,\cdots,B_s)$, a contradiction. Therefore, $\operatorname{span}(B_1,\cdots,B_s)=\mathbb{F}^n$ after all and so $\{B_1,\cdots,B_s\}$ is a basis for $\mathbb{F}^n$.

Now consider the block matrix,

$$S=\begin{pmatrix} B_1 & \cdots & B_s \end{pmatrix}$$

where

$$B_k=\begin{pmatrix} \mathbf{v}_1^k & \cdots & \mathbf{v}_{r_k}^k \end{pmatrix}.$$

Thus

$$S^{-1}=\begin{pmatrix} C_1 \\ \vdots \\ C_s \end{pmatrix}$$

where $C_iB_i=I_{r_i\times r_i}$ and $C_iNB_j=0$ if $i\neq j$. Let

$$C_k=\begin{pmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_{r_k}^T \end{pmatrix}.$$

Then

$$\begin{aligned} C_kNB_k &= \begin{pmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_{r_k}^T \end{pmatrix}\begin{pmatrix} N\mathbf{v}_1^k & \cdots & N\mathbf{v}_{r_k}^k \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_{r_k}^T \end{pmatrix}\begin{pmatrix} \mathbf{0} & \mathbf{v}_1^k & \cdots & \mathbf{v}_{r_k-1}^k \end{pmatrix} \end{aligned}$$

which equals an $r_k \times r_k$ matrix of the form

$$J_{r_k}(0) = \begin{pmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 1 \\ 0 & \cdots & \cdots & 0 \end{pmatrix}$$

That is, it has ones down the super diagonal and zeros everywhere else. It follows

$$\begin{aligned} S^{-1}NS &= \begin{pmatrix} C_1 \\ \vdots \\ C_s \end{pmatrix} N \begin{pmatrix} B_1 & \cdots & B_s \end{pmatrix} \\ &= \begin{pmatrix} J_{r_1}(0) & & & 0 \\ & J_{r_2}(0) & & \\ & & \ddots & \\ 0 & & & J_{r_s}(0) \end{pmatrix} \end{aligned}$$

as claimed. This proves the lemma.

Now let the upper triangular matrices, T_k be given in the conclusion of Corollary 15.4.4. Thus, as noted earlier,

$$T_k = \lambda_k I_{r_k \times r_k} + N_k$$

where N_k is a strictly upper triangular matrix of the sort just discussed in Lemma A.0.25. Therefore, there exists S_k such that $S_k^{-1}N_kS_k$ is of the form given in Lemma A.0.25. Now $S_k^{-1}\lambda_k I_{r_k \times r_k} S_k = \lambda_k I_{r_k \times r_k}$ and so $S_k^{-1}T_kS_k$ is of the form

$$\begin{pmatrix} J_{i_1}(\lambda_k) & & & 0 \\ & J_{i_2}(\lambda_k) & & \\ & & \ddots & \\ 0 & & & J_{i_s}(\lambda_k) \end{pmatrix}$$

where $i_1 \geq i_2 \geq \cdots \geq i_s$ and $\sum_{j=1}^{s} i_j = r_k$. This proves the following corollary.

Corollary A.0.26 *Suppose A is an upper triangular $n \times n$ matrix having α in every position on the main diagonal. Then there exists an invertible matrix, S such that*

$$S^{-1}AS = \begin{pmatrix} J_{k_1}(\alpha) & & & 0 \\ & J_{k_2}(\alpha) & & \\ & & \ddots & \\ 0 & & & J_{k_r}(\alpha) \end{pmatrix}$$

where $k_1 \geq k_2 \geq \cdots \geq k_r \geq 1$ and $\sum_{i=1}^{r} k_i = n$.

The next theorem is gives the existence of the Jordan canonical form.

Theorem A.0.27 *Let A be an $n \times n$ matrix having eigenvalues $\lambda_1, \cdots, \lambda_r$ where the multiplicity of λ_i as a zero of the characteristic polynomial equals m_i. Then there exists an invertible matrix, S such that*

$$S^{-1}AS = \begin{pmatrix} J(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & J(\lambda_r) \end{pmatrix} \tag{1.2}$$

where $J(\lambda_k)$ is an $m_k \times m_k$ matrix of the form

$$\begin{pmatrix} J_{k_1}(\lambda_k) & & & 0 \\ & J_{k_2}(\lambda_k) & & \\ & & \ddots & \\ 0 & & & J_{k_r}(\lambda_k) \end{pmatrix} \tag{1.3}$$

where $k_1 \geq k_2 \geq \cdots \geq k_r \geq 1$ and $\sum_{i=1}^{r} k_i = m_k$.

Proof: From Corollary 15.4.4, there exists S such that $S^{-1}AS$ is of the form

$$T \equiv \begin{pmatrix} T_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & T_r \end{pmatrix}$$

where T_k is an upper triangular $m_k \times m_k$ matrix having only λ_k on the main diagonal. By Corollary A.0.26 There exist matrices, S_k such that $S_k^{-1}T_kS_k = J(\lambda_k)$ where $J(\lambda_k)$ is described in 1.3. Now let M be the block diagonal matrix given by

$$M = \begin{pmatrix} S_1 & & 0 \\ & \ddots & \\ 0 & & S_r \end{pmatrix}.$$

It follows that $M^{-1}S^{-1}ASM = M^{-1}TM$ and this is of the desired form. This proves the theorem.

What about the uniqueness of the Jordan canonical form? Obviously if you change the order of the eigenvalues, you get a different Jordan canonical form but it turns out that if the order of the eigenvalues is the same, then the Jordan canonical form is unique. In fact, it is the same for any two similar matrices.

Theorem A.0.28 *Let A and B be two similar matrices. Let J_A and J_B be Jordan forms of A and B respectively, made up of the blocks $J_A(\lambda_i)$ and $J_B(\lambda_i)$ respectively. Then J_A and J_B are identical except possibly for the order of the $J(\lambda_i)$ where the λ_i are defined above.*

Proof: First note that for λ_i an eigenvalue, the matrices $J_A(\lambda_i)$ and $J_B(\lambda_i)$ are both of size $m_i \times m_i$ because the two matrices A and B, being similar, have exactly the same characteristic equation and the size of a block equals the algebraic multiplicity of the eigenvalue as a zero of the characteristic equation. It is only necessary to worry about the number and size of the Jordan blocks making up $J_A(\lambda_i)$ and $J_B(\lambda_i)$. Let the eigenvalues of A and B be $\{\lambda_1, \cdots, \lambda_r\}$. Consider the two sequences of numbers $\{\operatorname{rank}(A - \lambda I)^m\}$ and $\{\operatorname{rank}(B - \lambda I)^m\}$. Since A and B are similar, these two sequences coincide. (Why?) Also, for the same reason, $\{\operatorname{rank}(J_A - \lambda I)^m\}$ coincides with $\{\operatorname{rank}(J_B - \lambda I)^m\}$. Now pick λ_k an eigenvalue and consider $\{\operatorname{rank}(J_A - \lambda_k I)^m\}$ and $\{\operatorname{rank}(J_B - \lambda_k I)^m\}$. Then

$$J_A - \lambda_k I = \begin{pmatrix} J_A(\lambda_1 - \lambda_k) & & & & 0 \\ & \ddots & & & \\ & & J_A(0) & & \\ & & & \ddots & \\ 0 & & & & J_A(\lambda_r - \lambda_k) \end{pmatrix}$$

and a similar formula holds for $J_B - \lambda_k I$. Here

$$J_A(0) = \begin{pmatrix} J_{k_1}(0) & & & 0 \\ & J_{k_2}(0) & & \\ & & \ddots & \\ 0 & & & J_{k_r}(0) \end{pmatrix}$$

and

$$J_B(0) = \begin{pmatrix} J_{l_1}(0) & & & 0 \\ & J_{l_2}(0) & & \\ & & \ddots & \\ 0 & & & J_{l_p}(0) \end{pmatrix}$$

and it suffices to verify that $l_i = k_i$ for all i. As noted above, $\sum k_i = \sum l_i$. Now from the above formulas,

$$\begin{aligned} \operatorname{rank}(J_A - \lambda_k I)^m &= \sum_{i \neq k} m_i + \operatorname{rank}(J_A(0)^m) \\ &= \sum_{i \neq k} m_i + \operatorname{rank}(J_B(0)^m) \\ &= \operatorname{rank}(J_B - \lambda_k I)^m, \end{aligned}$$

which shows $\operatorname{rank}(J_A(0)^m) = \operatorname{rank}(J_B(0)^m)$ for all m. However,

$$J_B(0)^m = \begin{pmatrix} J_{l_1}(0)^m & & & 0 \\ & J_{l_2}(0)^m & & \\ & & \ddots & \\ 0 & & & J_{l_p}(0)^m \end{pmatrix}$$

with a similar formula holding for $J_A(0)^m$and $\operatorname{rank}(J_B(0)^m) = \sum_{i=1}^p \operatorname{rank}(J_{l_i}(0)^m)$, similar for $\operatorname{rank}(J_A(0)^m)$. In going from m to $m+1$,

$$\operatorname{rank}(J_{l_i}(0)^m) - 1 = \operatorname{rank}\left(J_{l_i}(0)^{m+1}\right)$$

untill $m = l_i$ at which time there is no further change. Therefore, $p = r$ since otherwise, there would exist a discrepancy right away in going from $m = 1$ to $m = 2$. Now suppose the sequence $\{l_i\}$ is not equal to the sequence, $\{k_i\}$. Then $l_{r-b} \neq k_{r-b}$ for some b a nonnegative integer taken to be a small as possible. Say $l_{r-b} > k_{r-b}$. Then, letting $m = k_{r-b}$,

$$\sum_{i=1}^r \operatorname{rank}(J_{l_i}(0)^m) = \sum_{i=1}^r \operatorname{rank}(J_{k_i}(0)^m)$$

and in going to $m+1$ a discrepancy must occur because the sum on the right will contribute less to the decrease in rank than the sum on the left. This proves the theorem.

An Assortment Of Worked Exercises And Examples

B.1 Worked Exercises

1. Here is an augmented matrix in which $*$ denotes an arbitrary number and $\blacksquare$ denotes a nonzero number. Determine whether the given augmented matrix is consistent. If consistent, is the solution unique?

$$\left(\begin{array}{ccccc|c} \blacksquare & * & * & * & * & * \\ 0 & \blacksquare & * & * & 0 & * \\ 0 & 0 & \blacksquare & * & * & \blacksquare \\ 0 & 0 & 0 & 0 & \blacksquare & * \end{array}\right)$$

   In this case the system is consistent and there is an infinite set of solutions. To see it is consistent, the bottom equation would yield a unique solution for x_5. Then letting $x_4 = t$, and substituting in to the other equations, beginning with the equation determined by the third row and then proceeding up to the next row followed by the first row, you get a solution for each value of t. There is a free variable which comes from the fourth column which is why you can say $x_4 = t$. Therefore, the solution is infinite.

2. Here is an augmented matrix in which $*$ denotes an arbitrary number and $\blacksquare$ denotes a nonzero number. Determine whether the given augmented matrix is consistent. If consistent, is the solution unique?

$$\left(\begin{array}{ccc|c} \blacksquare & * & * & * \\ 0 & 0 & \blacksquare & \blacksquare \\ 0 & 0 & * & 0 \end{array}\right)$$

   In this case there is no solution because you could use a row operation to place a 0 in the third row and third column position, like this:

$$\left(\begin{array}{ccc|c} \blacksquare & * & * & * \\ 0 & 0 & \blacksquare & \blacksquare \\ 0 & 0 & 0 & \blacksquare \end{array}\right)$$

   This would give a row of zeros equal to something nonzero.

3. Find h such that

$$\left(\begin{array}{cc|c} 1 & h & 4 \\ 3 & 7 & 7 \end{array}\right)$$

is the augmented matrix of an inconsistent matrix.

Doing a row operation by taking -3 times the top row and adding to the bottom, this gives

$$\begin{pmatrix} 1 & h & | & 4 \\ 0 & 7-3h & | & 7-12 \end{pmatrix}.$$

The system will be inconsistent if $7-3h=0$ or in other words, $h=7/3$.

4. Determine if the system is consistent.

$$\begin{aligned} x+2y+3z-w &= 2 \\ x-y+2z+w &= 1 \\ 2x+3y-z &= 1 \\ 4x+2y+z &= 5 \end{aligned}$$

The augmented matrix is

$$\begin{pmatrix} 1 & 2 & 3 & -1 & | & 2 \\ 1 & -1 & 2 & 1 & | & 1 \\ 2 & 3 & -1 & 0 & | & 1 \\ 4 & 2 & 1 & 0 & | & 5 \end{pmatrix}$$

A reduced echelon form for this is

$$\begin{pmatrix} 9 & 0 & 0 & 0 & | & 14 \\ 0 & 9 & 0 & 0 & | & -6 \\ 0 & 0 & 9 & 0 & | & 1 \\ 0 & 0 & 0 & 9 & | & -13 \end{pmatrix}.$$

Therefore, there is a unique solution. In particular the system is consistent.

5. Find the point, (x_1, y_1) which lies on both lines, $5x+3y=1$ and $4x-y=3$.

You solve the system of equations whose augmented matrix is

$$\begin{pmatrix} 5 & 3 & | & 1 \\ 4 & -1 & | & 3 \end{pmatrix}$$

A reduced echelon form is

$$\begin{pmatrix} 17 & 0 & 10 \\ 0 & 17 & -11 \end{pmatrix}$$

and so the solution is $x=17/10$ and $y=-11/17$.

6. Do the three lines, $3x+2y=1, 2x-y=1$, and $4x+3y=3$ have a common point of intersection? If so, find the point and if not, tell why they don't have such a common point of intersection.

This is asking for the solution to the three equations shown. The augmented matrix is

$$\begin{pmatrix} 3 & 2 & | & 1 \\ 2 & -1 & | & 1 \\ 4 & 3 & | & 3 \end{pmatrix}$$

A reduced echelon form is

$$\begin{pmatrix} 1 & 0 & | & 0 \\ 0 & 1 & | & 0 \\ 0 & 0 & | & 1 \end{pmatrix}$$

and this would require $0x+0y=1$ which is impossible so there is no solution to this system of equations and hence no point on each of the three lines.

7. Find the general solution of the system whose augmented matrix is

$$\left(\begin{array}{ccc|c} 1 & 2 & 0 & 2 \\ 1 & 1 & 4 & 2 \\ 2 & 3 & 4 & 4 \end{array}\right).$$

A reduced echelon form for the matrix is

$$\left(\begin{array}{cccc} 1 & 0 & 8 & 2 \\ 0 & 1 & -4 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

Therefore, $y = 4z$ and $x = 2 - 8z$. Apparently z can equal anything so we let $z = t$ and then the solution is

$$x = 2 - 8t, y = 4t, z = t.$$

8. Find the point, (x_1, y_1) which lies on both lines, $x + 2y = 1$ and $3x - y = 3$.

The solution is $y = 0$ and $x = 1$.

9. Find the point of intersection of the two lines $x + y = 3$ and $x + 2y = 1$.

The solution is $(5, -2)$.

10. Do the three lines, $x + 2y = 1, 2x - y = 1$, and $4x + 3y = 3$ have a common point of intersection? If so, find the point and if not, tell why they don't have such a common point of intersection.

To solve this set up the augmented matrix and go to work on it. The augmented matrix is

$$\left(\begin{array}{cc|c} 1 & 2 & 1 \\ 2 & -1 & 1 \\ 4 & 3 & 3 \end{array}\right)$$

A reduced echelon matrix for this is

$$\left(\begin{array}{cc|c} 1 & 0 & \frac{3}{5} \\ 0 & 1 & \frac{1}{5} \\ 0 & 0 & 0 \end{array}\right)$$

Therefore, there is a point in the intersection of these and it is $y = 1/5$ and $x = 3/5$. Thus the point is $(3/5, 1/5)$.

11. Do the three planes, $x + 2y - 3z = 2$, $x + y + z = 1$, and $3x + 2y + 2z = 0$ have a common point of intersection? If so, find one and if not, tell why there is no such point.

You need to find (x, y, z) which solves each equation. The augmented matrix is

$$\left(\begin{array}{ccc|c} 1 & 2 & -3 & 2 \\ 1 & 1 & 1 & 1 \\ 3 & 2 & 2 & 0 \end{array}\right)$$

A reduced echelon form for the matrix is

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & \frac{13}{5} \\ 0 & 0 & 1 & \frac{2}{5} \end{array}\right)$$

and so you should let $(x, y, z) = (-2, 13/5, 2/5)$.

12. Here is an augmented matrix in which $*$ denotes an arbitrary number and $\blacksquare$ denotes a nonzero number. Determine whether the given augmented matrix is consistent. If consistent, is the solution unique?

$$\left(\begin{array}{ccccc|c} \blacksquare & * & * & * & * & * \\ 0 & \blacksquare & * & * & 0 & * \\ 0 & 0 & \blacksquare & * & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * \end{array}\right)$$

You could do another set of row operations and reduce the matrix to one of the form

$$\left(\begin{array}{ccccc|c} \blacksquare & * & * & * & 0 & * \\ 0 & \blacksquare & * & * & 0 & * \\ 0 & 0 & \blacksquare & * & 0 & * \\ 0 & 0 & 0 & 0 & \blacksquare & * \end{array}\right)$$

It follows there exists a solution but the solution is not unique because x_4 is a free variable. You can pick it to be anything you like and the system will yield values for the other variables.

13. Here is an augmented matrix in which $*$ denotes an arbitrary number and $\blacksquare$ denotes a nonzero number. Determine whether the given augmented matrix is consistent. If consistent, is the solution unique?

$$\left(\begin{array}{ccc|c} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & \blacksquare & * \end{array}\right)$$

In this case there is a unique solution to the system. To see this, you could do more row operations and reduce this to something of the form

$$\left(\begin{array}{ccc|c} \blacksquare & 0 & 0 & * \\ 0 & \blacksquare & 0 & * \\ 0 & 0 & \blacksquare & * \end{array}\right).$$

14. Here is an augmented matrix in which $*$ denotes an arbitrary number and $\blacksquare$ denotes a nonzero number. Determine whether the given augmented matrix is consistent. If consistent, is the solution unique?

$$\left(\begin{array}{ccccc|c} \blacksquare & * & * & * & * & * \\ 0 & \blacksquare & 0 & * & 0 & * \\ 0 & 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 & \blacksquare & * \end{array}\right)$$

In this case, you could do more row operations and get something of the form

$$\left(\begin{array}{ccccc|c} \blacksquare & 0 & * & 0 & 0 & * \\ 0 & \blacksquare & 0 & 0 & 0 & * \\ 0 & 0 & 0 & \blacksquare & 0 & * \\ 0 & 0 & 0 & 0 & \blacksquare & * \end{array}\right)$$

Now you can determine the answer.

15. Find h such that
$$\left(\begin{array}{cc|c} 2 & h & 4 \\ 3 & 6 & 7 \end{array}\right)$$
is the augmented matrix of an inconsistent matrix.

Take -3 times the top row and add to 2 times the bottom. This yields
$$\left(\begin{array}{cc|c} 2 & h & 4 \\ 0 & 12-3h & 2 \end{array}\right)$$
Now if $h = 4$ the system is inconsistent because it would have the bottom row equal to $\left(\begin{array}{cc|c} 0 & 0 & 2 \end{array}\right)$.

16. Choose h and k such that the augmented matrix shown has one solution. Then choose h and k such that the system has no solutions. Finally, choose h and k such that the system has infinitely many solutions.
$$\left(\begin{array}{cc|c} 1 & h & 2 \\ 2 & 4 & k \end{array}\right).$$

If $h \neq 2$ then k can be anything and the system represented by the augmented matrix will have a unique solution. Suppose then that $h = 2$. Then taking -2 times the top row and adding to the bottom row gives
$$\left(\begin{array}{cc|c} 1 & 2 & 2 \\ 0 & 0 & k-4 \end{array}\right)$$
If $k \neq 4$ there is no solution. However, if $k = 4$ you are left with the single equation, $x+2y = 2$ and there are infinitely many solutions to this. In fact anything of the form $(2-2y, y)$ will work just fine.

17. Determine if the system is consistent.
$$\begin{aligned} x + 2y + z - w &= 2 \\ x - y + z + w &= 1 \\ 2x + y - z &= 1 \\ 4x + 2y + z &= 5 \end{aligned}$$

This system is inconsistent. To see this, write the augmented matrix and do row operations. The augmented matrix is
$$\left(\begin{array}{cccc|c} 1 & 2 & 1 & -1 & 2 \\ 1 & -1 & 1 & 1 & 1 \\ 2 & 1 & -1 & 0 & 1 \\ 4 & 2 & 1 & 0 & 5 \end{array}\right)$$
A reduced echelon form for this matrix is
$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & \frac{1}{3} & 0 \\ 0 & 1 & 0 & -\frac{2}{3} & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right)$$
and the bottom row shows there is no solution.

18. Find the general solution of the system whose augmented matrix is

$$\left(\begin{array}{ccc|c} 1 & 2 & 0 & 2 \\ 1 & 3 & 4 & 2 \\ 1 & 0 & 2 & 1 \end{array}\right)$$

A reduced echelon form for this matrix is

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & \frac{6}{5} \\ 0 & 1 & 0 & \frac{2}{5} \\ 0 & 0 & 1 & -\frac{1}{10} \end{array}\right)$$

and so the solution is unique and is $z = -1/10, y = 2/5$, and $x = 6/5$.

19. Find the general solution of the system whose augmented matrix is

$$\left(\begin{array}{ccc|c} 1 & 1 & 0 & 5 \\ 1 & 0 & 3 & 2 \end{array}\right).$$

A reduced echelon form for this matrix is

$$\left(\begin{array}{ccc|c} 1 & 0 & 3 & 2 \\ 0 & 1 & -3 & 3 \end{array}\right)$$

and so the general solution is of the form $y = 3 + 3z, x = 2 - 3z$ with z arbitrary.

20. Find the general solution of the system whose augmented matrix is

$$\left(\begin{array}{ccccc|c} 1 & 0 & 2 & 1 & 1 & 3 \\ 0 & 1 & 0 & 4 & 2 & 1 \\ 2 & 2 & 0 & 0 & 1 & 3 \\ 1 & 0 & 1 & 0 & 2 & 2 \end{array}\right).$$

You do the usual thing, row operations on the matrix to obtain a reduced echelon form. A reduced echelon form is

$$\left(\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & \frac{9}{2} & \frac{7}{6} \\ 0 & 1 & 0 & 0 & -4 & \frac{1}{3} \\ 0 & 0 & 1 & 0 & -\frac{5}{2} & \frac{5}{6} \\ 0 & 0 & 0 & 1 & \frac{3}{2} & \frac{1}{6} \end{array}\right)$$

Therefore, the general solution is $x_4 = 1/6 - 3/2x_5, x_3 = 5/6 + 5/2x_5, x_2 = 1/3 + 4x_5$, and $x_1 = 7/6 - 9/2x_5$ with x_5 arbitrary.

B.2 Worked Exercises

1. Here are some matrices:

$$\begin{aligned} A &= \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 7 \\ 1 & 0 & 1 \end{pmatrix}, B = \begin{pmatrix} 3 & -1 & 2 \\ -3 & 2 & 1 \end{pmatrix}, \\ C &= \begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 1 & 1 \end{pmatrix}, D = \begin{pmatrix} -1 & 2 \\ 2 & -3 \end{pmatrix}, E = \begin{pmatrix} 2 \\ 3 \end{pmatrix}. \end{aligned}$$

Find if possible $-3A, 3B-A, AC, CB, EA, DC^T$ If it is not possible explain why.

$$-3A = -3\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 7 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -3 & -6 & -9 \\ -6 & -9 & -21 \\ -3 & 0 & -3 \end{pmatrix}$$

$3B-A$ is nonsense because the matrices B and A are not of the same size.

$$AC = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 7 \\ 1 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 10 & 7 \\ 18 & 14 \\ 2 & 3 \end{pmatrix}$$

There is no problem here because you are doing $(3\times 3)(3\times 2)$.

$$CB = \begin{pmatrix} 1 & 2 \\ 3 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 3 & -1 & 2 \\ -3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} -3 & 3 & 4 \\ 6 & -1 & 7 \\ 0 & 1 & 3 \end{pmatrix}$$

There is no problem here because you are doing $(3\times 2)(2\times 3)$ and the inside numbers match. EA is nonsense because it is of the form $(2\times 1)(3\times 3)$ so since the inside numbers do not match the matrices are not conformable.

$$DC^T = \begin{pmatrix} -1 & 2 \\ 2 & -3 \end{pmatrix}\begin{pmatrix} 1 & 3 & 1 \\ 2 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & -1 & 1 \\ -4 & 3 & -1 \end{pmatrix}.$$

2. Let $A = \begin{pmatrix} 0 & 2 \\ 3 & 4 \end{pmatrix}, B = \begin{pmatrix} 1 & 2 \\ 1 & k \end{pmatrix}$. Is it possible to choose k such that $AB = BA$? If so, what should k equal?

   We just multiply and see if it can happen.

$$AB = \begin{pmatrix} 0 & 2 \\ 3 & 4 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 1 & k \end{pmatrix} = \begin{pmatrix} 2 & 2k \\ 7 & 6+4k \end{pmatrix}.$$

   On the other hand,

$$BA = \begin{pmatrix} 1 & 2 \\ 1 & k \end{pmatrix}\begin{pmatrix} 0 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 6 & 10 \\ 3k & 2+4k \end{pmatrix}.$$

   If these were equal you would need to have $6 = 2$ which is not the case. Therefore, there is no way to choose k such that these two matrices will commute.

3. Let $\mathbf{x} = (-1, 0, 3)$ and $\mathbf{y} = (3, 1, 2)$. Find $\mathbf{x}^T\mathbf{y}$.

$$\mathbf{x}^T\mathbf{y} = \begin{pmatrix} -1 \\ 0 \\ 3 \end{pmatrix}\begin{pmatrix} 3 & 1 & 2 \end{pmatrix} = \begin{pmatrix} -3 & -1 & -2 \\ 0 & 0 & 0 \\ 9 & 3 & 6 \end{pmatrix}.$$

4. Write $\begin{pmatrix} 4x_1 - x_2 + 2x_3 \\ 2x_3 + 7x_1 \\ 2x_3 \\ 3x_3 + 3x_2 + x_1 \end{pmatrix}$ in the form $A\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$ where A is an appropriate matrix.

$$\begin{pmatrix} 4 & -1 & 2 & 0 \\ 7 & 0 & 2 & 0 \\ 0 & 0 & 2 & 0 \\ 1 & 3 & 3 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}.$$

5. Let
$$A = \begin{pmatrix} 1 & 2 & 5 \\ 2 & 1 & 4 \\ 1 & 0 & 2 \end{pmatrix}.$$
Find A^{-1} if possible. If A^{-1} does not exist, determine why.
$$\begin{pmatrix} 1 & 2 & 5 \\ 2 & 1 & 4 \\ 1 & 0 & 2 \end{pmatrix}^{-1} = \begin{pmatrix} -\frac{2}{3} & \frac{4}{3} & -1 \\ 0 & 1 & -2 \\ \frac{1}{3} & -\frac{2}{3} & 1 \end{pmatrix}.$$

6. Let
$$A = \begin{pmatrix} 1 & 2 & 0 & 2 \\ 1 & 5 & 2 & 0 \\ 2 & 1 & -3 & 2 \\ 1 & 2 & 1 & 2 \end{pmatrix}$$
Find A^{-1} if possible. If A^{-1} does not exist, determine why.
$$\begin{pmatrix} 1 & 2 & 0 & 2 \\ 1 & 5 & 2 & 0 \\ 2 & 1 & -3 & 2 \\ 1 & 2 & 1 & 2 \end{pmatrix}^{-1} = \begin{pmatrix} -3 & \frac{1}{6} & \frac{5}{6} & \frac{13}{6} \\ 1 & \frac{1}{6} & -\frac{1}{6} & -\frac{5}{6} \\ -1 & 0 & 0 & 1 \\ 1 & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \end{pmatrix}.$$

7. Show that if A^{-1} exists for an $n \times n$ matrix, then it is unique. That is, if $BA = I$ and $AB = I$, then $B = A^{-1}$.

   From $AB = I$, multiply both sides by A^{-1}. Thus $A^{-1}(AB) = A^{-1}$. Then from the associative property of matrix multiplication, $A^{-1} = A^{-1}(AB) = \left(A^{-1}A\right)B = IB = B$.

B.3 Worked Exercises

1. Find the following determinant by expanding along the second column.
$$\begin{vmatrix} 1 & 3 & 1 \\ 2 & 1 & 5 \\ 2 & 1 & 1 \end{vmatrix}$$
   This is
$$3(-1)^{2+1}\begin{vmatrix} 2 & 5 \\ 2 & 1 \end{vmatrix} + 1(-1)^{1+1}\begin{vmatrix} 1 & 1 \\ 2 & 1 \end{vmatrix} + 1(-1)^{3+2}\begin{vmatrix} 1 & 1 \\ 2 & 5 \end{vmatrix} = 20.$$

2. Compute the determinant by cofactor expansion. Pick the easiest row or column to use.
$$\begin{vmatrix} 2 & 0 & 0 & 1 \\ 2 & 1 & 1 & 0 \\ 0 & 0 & 0 & 3 \\ 2 & 3 & 3 & 1 \end{vmatrix}$$
   You ought to use the third row. This yields the above equals
$$3\begin{vmatrix} 2 & 0 & 0 \\ 2 & 1 & 1 \\ 2 & 3 & 3 \end{vmatrix} = (3)(2)\begin{vmatrix} 1 & 1 \\ 3 & 3 \end{vmatrix} = 0.$$

3. Find the determinant using row and column operations.

$$\begin{vmatrix} 5 & 4 & 3 & 2 \\ 3 & 2 & 4 & 3 \\ -1 & 2 & 3 & 3 \\ 2 & 1 & 2 & -2 \end{vmatrix}$$

Replace the first row by 5 times the third added to it and then replace the second by 3 times the third added to it and then the last by 2 times the third added to it. This yields

$$\begin{vmatrix} 0 & 14 & 18 & 17 \\ 0 & 8 & 13 & 12 \\ -1 & 2 & 3 & 3 \\ 0 & 5 & 8 & 4 \end{vmatrix}$$

Now lets replace the third column by -1 times the last column added to it.

$$\begin{vmatrix} 0 & 14 & 1 & 17 \\ 0 & 8 & 1 & 12 \\ -1 & 2 & 0 & 3 \\ 0 & 5 & 4 & 4 \end{vmatrix}$$

Now replace the top row by -1 times the second added to it and the bottom row by -4 times the second added to it. This yields

$$\begin{vmatrix} 0 & 6 & 0 & 5 \\ 0 & 8 & 1 & 12 \\ -1 & 2 & 0 & 3 \\ 0 & -27 & 0 & -44 \end{vmatrix}. \tag{2.1}$$

This looks pretty good because it has a lot of zeros. Expand along the first column and next along the second,

$$(-1)\begin{vmatrix} 6 & 0 & 5 \\ 8 & 1 & 12 \\ -27 & 0 & -44 \end{vmatrix} = (-1)(1)\begin{vmatrix} 6 & 5 \\ -27 & -44 \end{vmatrix} = 129.$$

Alternatively, you could continue doing row and column operations. Switch the third and first row in 2.1 to obtain

$$-\begin{vmatrix} -1 & 2 & 0 & 3 \\ 0 & 8 & 1 & 12 \\ 0 & 6 & 0 & 5 \\ 0 & -27 & 0 & -44 \end{vmatrix}$$

Next take $9/2$ times the third row and add to the bottom.

$$-\begin{vmatrix} -1 & 2 & 0 & 3 \\ 0 & 8 & 1 & 12 \\ 0 & 6 & 0 & 5 \\ 0 & 0 & 0 & -44+(9/2)5 \end{vmatrix}.$$

Finally, take $-6/8$ times the second row and add to the third.

$$-\begin{vmatrix} -1 & 2 & 0 & 3 \\ 0 & 8 & 1 & 12 \\ 0 & 0 & -6/8 & 5+(-6/8)(12) \\ 0 & 0 & 0 & -44+(9/2)5 \end{vmatrix}.$$

Therefore, since the matrix is now upper triangular, the determinant is

$$-\left((-1)\left(8\right)\left(-6/8\right)\left(-44+\left(9/2\right)5\right)\right)=129.$$

4. An operation is done to get from the first matrix to the second. Identify what was done and tell how it will affect the value of the determinant.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

This involved taking the transpose so the determinant of the new matrix is the same as the determinant of the first matrix.

5. Show that for A a 2×2 matrix $\det\left(aA\right)=a^2\det\left(A\right)$ where a is a scalar.

$a^2\det\left(A\right)=a\det\left(A_1\right)$ where the first row of A is replaced by a times it to get A_1. Then $a\det\left(A_1\right)=A_2$ where A_2 is obtained from A by multiplying both rows by a. In other words, $A_2=aA$. Thus the conclusion is established.

6. Use Cramer's rule to find y in

$$\begin{array}{c} 2x+2y+z=3 \\ 2x-y-z=2 \\ x+2z=1 \end{array}$$

From Cramer's rule,

$$y=\frac{\begin{vmatrix} 2 & 3 & 1 \\ 2 & 2 & -1 \\ 1 & 1 & 2 \end{vmatrix}}{\begin{vmatrix} 2 & 2 & 1 \\ 2 & -1 & -1 \\ 1 & 0 & 2 \end{vmatrix}}=\frac{5}{13}.$$

7. Here is a matrix,

$$\begin{pmatrix} e^t & e^{-t}\cos t & e^{-t}\sin t \\ e^t & -e^{-t}\cos t-e^{-t}\sin t & -e^{-t}\sin t+e^{-t}\cos t \\ e^t & 2e^{-t}\sin t & -2e^{-t}\cos t \end{pmatrix}$$

Does there exist a value of t for which this matrix fails to have an inverse? Explain.

$\det\begin{pmatrix} e^t & e^{-t}\cos t & e^{-t}\sin t \\ e^t & -e^{-t}\cos t-e^{-t}\sin t & -e^{-t}\sin t+e^{-t}\cos t \\ e^t & 2e^{-t}\sin t & -2e^{-t}\cos t \end{pmatrix}$
$=5e^te^{2(-t)}\cos^2 t+5e^te^{2(-t)}\sin^2 t=5e^{-t}$ which is never equal to zero for any value of t and so there is no value of t for which the matrix has no inverse.

8. Use the formula for the inverse in terms of the cofactor matrix to find if possible the inverse of the matrix

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 6 & 1 \\ 4 & 1 & 1 \end{pmatrix}.$$

First you need to take the determinant

$$\det\begin{pmatrix} 1 & 2 & 3 \\ 0 & 6 & 1 \\ 4 & 1 & 1 \end{pmatrix}=-59$$

and so the matrix has an inverse. Now you need to find the cofactor matrix.

$$\begin{pmatrix} \begin{vmatrix} 6 & 1 \\ 1 & 1 \end{vmatrix} & -\begin{vmatrix} 0 & 1 \\ 4 & 1 \end{vmatrix} & \begin{vmatrix} 0 & 6 \\ 4 & 1 \end{vmatrix} \\ -\begin{vmatrix} 2 & 3 \\ 1 & 1 \end{vmatrix} & \begin{vmatrix} 1 & 3 \\ 4 & 1 \end{vmatrix} & -\begin{vmatrix} 1 & 2 \\ 4 & 1 \end{vmatrix} \\ \begin{vmatrix} 2 & 3 \\ 6 & 1 \end{vmatrix} & -\begin{vmatrix} 1 & 3 \\ 0 & 1 \end{vmatrix} & \begin{vmatrix} 1 & 2 \\ 0 & 6 \end{vmatrix} \end{pmatrix}$$

$$= \begin{pmatrix} 5 & 4 & -24 \\ 1 & -11 & 7 \\ -16 & -1 & 6 \end{pmatrix}.$$

Thus the inverse is

$$\frac{1}{-59}\begin{pmatrix} 5 & 4 & -24 \\ 1 & -11 & 7 \\ -16 & -1 & 6 \end{pmatrix}^T$$

$$= \frac{1}{-59}\begin{pmatrix} 5 & 1 & -16 \\ 4 & -11 & -1 \\ -24 & 7 & 6 \end{pmatrix}.$$

If you check this, it does work.

B.4 Worked Exercises

1. Find the rank of the following matrices. If the rank is r, identify r columns **in the original matrix** which have the property that every other column may be written as a linear combination of these. Also find a basis for the row and column spaces of the matrices.

   (a) $\begin{pmatrix} 9 & 2 & 0 \\ 3 & 7 & 1 \\ 6 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix}$

   From using row operations we obtain the row reduced echelon form which is

   $$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

   Therefore, a basis for the column space of the original matrix is the first three columns of the original matrix. A basis for the row space is just

   $$\begin{pmatrix} 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}, \text{and } \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}.$$

   (b) $\begin{pmatrix} 3 & 0 & 3 \\ 10 & 9 & 1 \\ 1 & 1 & 0 \\ 2 & 2 & 0 \end{pmatrix}$

In this case the row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and so a basis for the column space of the original matrix consists of the first two columns of the original matrix and a basis for the row space is $\begin{pmatrix} 1 & 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 0 & 1 & -1 \end{pmatrix}$.

(c) $\begin{pmatrix} 0 & 1 & 7 & 8 & 1 & 9 & 2 \\ 0 & 3 & 2 & 5 & 1 & 6 & 8 \\ 0 & 1 & 1 & 2 & 0 & 2 & 3 \\ 0 & 2 & 1 & 3 & 0 & 3 & 4 \end{pmatrix}$

The row reduced echelon form of this matrix is

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

and so a basis for the column space of the original matrix consists of the second, third, fifth, and seventh columns of the original matrix. A basis for the row space consists of the rows of this last matrix in row reduced echelon form.

2. Let H denote span $\left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 4 \\ 5 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right)$. Find the dimension of H and determine a basis.

Make these the columns of a matrix and ask for the rank of this matrix.

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 4 & 3 & 1 \\ 0 & 5 & 1 & 1 \end{pmatrix}$$

The row reduced echelon form is

$$\begin{pmatrix} 1 & 0 & 0 & \frac{8}{7} \\ 0 & 1 & 0 & \frac{2}{7} \\ 0 & 0 & 1 & -\frac{3}{7} \end{pmatrix}$$

A basis for H is

$$\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 4 \\ 5 \end{pmatrix}, \begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix} \right\}$$

and so H has dimension 3.

3. Here are three vectors. Determine whether they are linearly independent or linearly dependent.

$$\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix}$$

You need to consider the solutions to the equation

$$c_1\begin{pmatrix}1\\0\\1\end{pmatrix}+c_2\begin{pmatrix}2\\0\\1\end{pmatrix}+c_3\begin{pmatrix}3\\0\\0\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}$$

and determine whether there is a solution other than the obvious one, $c_1=c_2=c_3=0$. The augmented matrix for the system of equations is

$$\left(\begin{array}{ccc|c}1&2&3&0\\0&0&0&0\\1&1&0&0\end{array}\right)$$

Taking -1 times the top row and adding to the bottom and then switching the two bottom rows yields

$$\left(\begin{array}{ccc|c}1&2&3&0\\0&-1&-3&0\\0&0&0&0\end{array}\right)$$

Next take 2 times the second row and add to the top. This yields

$$\left(\begin{array}{ccc|c}1&0&-3&0\\0&-1&-3&0\\0&0&0&0\end{array}\right)$$

There are solutions other than the zero solution because c_3 is a free variable. Therefore, these vectors are not linearly independent.

4. Here are four vectors. Determine whether they span $\mathbb{R}^3$. Are these vectors linearly independent?

$$\begin{pmatrix}1\\2\\3\end{pmatrix},\begin{pmatrix}4\\0\\3\end{pmatrix},\begin{pmatrix}3\\2\\0\end{pmatrix},\begin{pmatrix}2\\1\\6\end{pmatrix}$$

The vectors can't possibly be linearly independent. If they were, they would constitute a linearly independent set consisting of four vectors even though there exists a spanning set of only three,

$$\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}0\\1\\0\end{pmatrix},\begin{pmatrix}0\\0\\1\end{pmatrix}$$

However, the four given vectors might still span $\mathbb{R}^3$ even though they are not a basis. What does it take to span $\mathbb{R}^3$? Given a vector $(x,y,z)^T\in\mathbb{R}^3$, do there exist scalars c_1,c_2,c_3, and c_4 such that

$$c_1\begin{pmatrix}1\\2\\3\end{pmatrix}+c_2\begin{pmatrix}4\\0\\3\end{pmatrix}+c_3\begin{pmatrix}3\\2\\0\end{pmatrix}+c_4\begin{pmatrix}2\\1\\6\end{pmatrix}=\begin{pmatrix}x\\y\\z\end{pmatrix}?$$

Consider the augmented matrix of the above,

$$\left(\begin{array}{cccc|c}1&4&3&2&x\\2&0&2&1&y\\3&3&0&6&z\end{array}\right)$$

Doing row operations till an echelon form is obtained leads to

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & \frac{5}{4} & \frac{1}{4}y+\frac{2}{9}z-\frac{1}{6}x \\ 0 & 1 & 0 & \frac{3}{4} & -\frac{1}{4}y+\frac{1}{6}x+\frac{1}{9}z \\ 0 & 0 & 1 & -\frac{3}{4} & -\frac{2}{9}z+\frac{1}{6}x+\frac{1}{4}y \end{array}\right)$$

and you see there is a solution to the desired system of equations. In fact there are infinitely many because c_4 is a free variable. Therefore, the four vectors do span $\mathbb{R}^3$.

5. Consider the vectors of the form

$$\left\{\left(\begin{array}{c} 2t+6s \\ s-2t \\ 3t+s \end{array}\right) : s,t \in \mathbb{R}\right\}.$$

Is this set of vectors a subspace of $\mathbb{R}^3$? If so, explain why, give a basis for the subspace and find its dimension.

This is indeed a subspace. You only need to verify the set of vectors is closed with respect to the vector space operations. Let $\left(\begin{array}{c} 2t_1+6s_1 \\ s_1-2t_1 \\ 3t_1+s_1 \end{array}\right)$ and $\left(\begin{array}{c} 2t+6s \\ s-2t \\ 3t+s \end{array}\right)$ be two vectors in the given set of vectors.

$$\begin{aligned} & \alpha\left(\begin{array}{c} 2t+6s \\ s-2t \\ 3t+s \end{array}\right)+\beta\left(\begin{array}{c} 2t_1+6s_1 \\ s_1-2t_1 \\ 3t_1+s_1 \end{array}\right) \\ = & \left(\begin{array}{c} 2\alpha t+6\alpha s+2\beta t_1+6\beta s_1 \\ \alpha s-2\alpha t+\beta s_1-2\beta t_1 \\ 3\alpha t+\alpha s+3\beta t_1+\beta s_1 \end{array}\right) \\ = & \left(\begin{array}{c} 2\left(\alpha t+\beta t_1\right)+6\left(\alpha s+\beta s_1\right) \\ \alpha s+\beta s_1-2\left(\alpha t+\beta t_1\right) \\ 3\left(\alpha t+\beta t_1\right)+\alpha s+\beta s_1 \end{array}\right) \end{aligned}$$

If we let $T \equiv \alpha t+\beta t_1$ and $S \equiv \alpha s+\beta s_1$, this is seen to be of the form

$$\left(\begin{array}{c} 2T+6S \\ S-2T \\ 3T+S \end{array}\right)$$

which is the way the vectors in the given set are described. Another way to see this is to notice that the vectors in the given set are of the form

$$t\left(\begin{array}{c} 2 \\ -2 \\ 3 \end{array}\right)+s\left(\begin{array}{c} 6 \\ 1 \\ 1 \end{array}\right)$$

so it consists of the span of the two vectors,

$$\left(\begin{array}{c} 2 \\ -2 \\ 3 \end{array}\right), \left(\begin{array}{c} 6 \\ 1 \\ 1 \end{array}\right). \tag{2.2}$$

Recall that the span of a set of vectors is always a subspace. You can also verify these vectors in 2.2 form a linearly independent set and so they are a basis.

6. Let $M = \left\{\mathbf{u} = (u_1, u_2, u_3, u_4) \in \mathbb{R}^4 : u_3 \geq u_2\right\}$. Is M a subspace? Explain.

   This is not a subspace because if $\mathbf{u} \in M$, is such that $u_3 > u_2$, then consider $(-1)\,\mathbf{u}$. If this were in M you would need to have $-u_3 > -u_2$ and so $u_3 < u_2$ which cannot be true if $u_3 > u_2$. Thus M is not closed under scalar multiplication so it is not a subspace.

7. Let $\mathbf{w}, \mathbf{w}_1$ be given vectors in $\mathbb{R}^2$ and define

$$M = \left\{\mathbf{u} = (u_1, u_2) \in \mathbb{R}^2 : \mathbf{w} \cdot \mathbf{u} = 0 \text{ and } \mathbf{w}_1 \cdot \mathbf{u} = 0\right\}.$$

   Is M a subspace? Explain.

   Suppose $\mathbf{u}'$ and $\mathbf{u}$ are both in M. What about $\alpha\mathbf{u}' + \beta\mathbf{u}$?

$$\mathbf{w}\cdot(\alpha\mathbf{u}' + \beta\mathbf{u}) = \alpha\mathbf{w}\cdot\mathbf{u}' + \beta\mathbf{w}\cdot\mathbf{u} = \alpha 0 + \beta 0 = 0$$

   Similarly,

$$\mathbf{w}_1\cdot(\alpha\mathbf{u}' + \beta\mathbf{u}) = \alpha\mathbf{w}_1\cdot\mathbf{u}' + \beta\mathbf{w}_1\cdot\mathbf{u} = \alpha 0 + \beta 0 = 0$$

   and so $\alpha\mathbf{u}' + \beta\mathbf{u} \in M$. This has verified that M is a subspace.

B.5 Worked Exercises

1. Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $5\pi/12$.

   You note that $5\pi/12 = 2\pi/3 - \pi/4$. Therefore, you can first rotate through $-\pi/4$ and then rotate through $2\pi/3$ to get the rotation through $5\pi/12$. The matrix of the transformation with respect to the usual coordinates which rotates through $-\pi/4$ is

$$\begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}$$

   and the matrix of the transformation which rotates through $2\pi/3$ is

$$\begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix}.$$

   Multiplying these gives

$$\begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix}\begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}$$

$$= \begin{pmatrix} -\frac{1}{4}\sqrt{2} + \frac{1}{4}\sqrt{3}\sqrt{2} & -\frac{1}{4}\sqrt{2} - \frac{1}{4}\sqrt{3}\sqrt{2} \\ \frac{1}{4}\sqrt{3}\sqrt{2} + \frac{1}{4}\sqrt{2} & -\frac{1}{4}\sqrt{2} + \frac{1}{4}\sqrt{3}\sqrt{2} \end{pmatrix}$$

   and this is the matrix of the desired transformation. Note this shows that

$$\cos(5\pi/12) = -\frac{1}{4}\sqrt{2} + \frac{1}{4}\sqrt{3}\sqrt{2} \approx .258\,819\,05$$

$$\sin(5\pi/12) = \frac{1}{4}\sqrt{3}\sqrt{2} + \frac{1}{4}\sqrt{2} \approx .965\,925\,83.$$

2. Find the matrix for the linear transformation which rotates every vector in $\mathbb{R}^2$ through an angle of $2\pi/3$ and then reflects across the x axis.

   What does it do to $\mathbf{e}_1$? First you rotate $\mathbf{e}_1$ through the given angle to obtain

   $$\begin{pmatrix} -1/2 \\ \sqrt{3}/2 \end{pmatrix}$$

   and then this becomes

   $$\begin{pmatrix} -1/2 \\ -\sqrt{3}/2 \end{pmatrix}.$$

   This is the first column of the desired matrix. Next $\mathbf{e}_2$ first is rotated through the given angle to give

   $$\begin{pmatrix} -\sqrt{3}/2 \\ -1/2 \end{pmatrix}$$

   and then it is reflected across the x axis to give

   $$\begin{pmatrix} -\sqrt{3}/2 \\ 1/2 \end{pmatrix}$$

   and this gives the second column of the desired matrix. Thus the matrix is

   $$\begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ -\sqrt{3}/2 & 1/2 \end{pmatrix}.$$

3. Find the matrix for $\mathrm{proj}_{\mathbf{u}}(\mathbf{v})$ where $\mathbf{u} = (1,-2,3)^T$.

   Recall

   $$\mathrm{proj}_{\mathbf{u}}(\mathbf{v}) = \frac{\mathbf{v}\cdot\mathbf{u}}{||\mathbf{u}||^2}\mathbf{u}$$

   Therefore,

   $$\begin{aligned} \mathrm{proj}_{\mathbf{u}}(\mathbf{e}_1) &= \frac{1}{14}\begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix},\ \mathrm{proj}_{\mathbf{u}}(\mathbf{e}_2) = \frac{-2}{14}\begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}, \\ \mathrm{proj}_{\mathbf{u}}(\mathbf{e}_2) &= \frac{3}{14}\begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}. \end{aligned}$$

   Hence the desired matrix is

   $$\frac{1}{14}\begin{pmatrix} 1 & -2 & 3 \\ -2 & 4 & -6 \\ 3 & -6 & 9 \end{pmatrix}.$$

4. Show that the function $T_{\mathbf{u}}$ defined by $T_{\mathbf{u}}(\mathbf{v}) \equiv \mathbf{v} - \mathrm{proj}_{\mathbf{u}}(\mathbf{v})$ is also a linear transformation.

   $$T_{\mathbf{u}}(\alpha\mathbf{v}+\beta\mathbf{w}) = \alpha\mathbf{v}+\beta\mathbf{w} - \mathrm{proj}_{\mathbf{u}}(\alpha\mathbf{v}+\beta\mathbf{w})$$

   which from 3 equals

   $$\alpha(\mathbf{v}-\mathrm{proj}_{\mathbf{u}}(\mathbf{v})) + \beta(\mathbf{w}-\mathrm{proj}_{\mathbf{u}}(\mathbf{w})) = \alpha T_{\mathbf{u}}\mathbf{v} + \beta T_{\mathbf{u}}\mathbf{w}.$$

   This is what it takes to be a linear transformation.

5. If A, B, and C are each $n \times n$ matrices and ABC is invertible, why are each of A, B, and C invertible.

   $0 \neq \det(ABC) = \det(A)\det(B)\det(C)$ and so none of $\det(A), \det(B)$, or $\det(C)$ can equal zero. Therefore, each is invertible. You should do this another way, showing that each of A, B, and C is one to one and then using a theorem presented earlier.

6. Give an example of a 3×1 matrix with the property that the linear transformation determined by this matrix is one to one but not onto.

   Here is one. $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. If $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} x = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$, then $x = 0$ but this is certainly not onto as a map from $\mathbb{R}^1$ to $\mathbb{R}^3$ because it does not ever yield $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$.

7. Find the matrix of the linear transformation from $\mathbb{R}^3$ to $\mathbb{R}^3$ which first rotates every vector through an angle of $\pi/4$ about the z axis when viewed from the positive z axis and then rotates every vector through an angle of $\pi/6$ about the x axis when viewed from the positive x axis.

   The matrix of the linear transformation which accomplishes the first rotation is

$$\begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 & 0 \\ \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

   and the matrix which accomplishes the second rotation is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \sqrt{3}/2 & -1/2 \\ 0 & 1/2 & \sqrt{3}/2 \end{pmatrix}$$

   Therefore, the matrix of the desired linear transformation is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \sqrt{3}/2 & -1/2 \\ 0 & 1/2 & \sqrt{3}/2 \end{pmatrix} \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 & 0 \\ \sqrt{2}/2 & \sqrt{2}/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{2}\sqrt{2} & -\frac{1}{2}\sqrt{2} & 0 \\ \frac{1}{4}\sqrt{3}\sqrt{2} & \frac{1}{4}\sqrt{3}\sqrt{2} & -\frac{1}{2} \\ \frac{1}{4}\sqrt{2} & \frac{1}{4}\sqrt{2} & \frac{1}{2}\sqrt{3} \end{pmatrix}$$

   This might not be the first thing you would think of.

B.6 Worked Exercises

1. Find an LU factorization of $\begin{pmatrix} 1 & 2 & 7 \\ 3 & 1 & 3 \\ 1 & 2 & 3 \end{pmatrix}$.

   To find this we write

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 7 \\ 3 & 1 & 3 \\ 1 & 2 & 3 \end{pmatrix}$$

and put the one on the right into echelon form and keep track of the multipliers. Updating the first column,

$$\begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 7 \\ 0 & -5 & -18 \\ 0 & 0 & -4 \end{pmatrix}$$

We now stop because the matrix on the right is upper triangular.

2. Find an LU factorization of $\begin{pmatrix} 1 & 7 & 3 & 2 \\ 1 & 3 & 8 & 1 \\ 5 & 1 & 1 & 3 \end{pmatrix}$.

To find it we write

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 7 & 3 & 2 \\ 1 & 3 & 8 & 1 \\ 5 & 1 & 1 & 3 \end{pmatrix}$$

and update keeping track of the multipliers. First we update the first column.

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 5 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 7 & 3 & 2 \\ 0 & -4 & 5 & -1 \\ 0 & -34 & -14 & -7 \end{pmatrix}$$

Next we update the second column.

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 5 & 34/4 & 1 \end{pmatrix} \begin{pmatrix} 1 & 7 & 3 & 2 \\ 0 & -4 & 5 & -1 \\ 0 & 0 & -14-34/4\times 5 & -7+34/4 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 5 & 34/4 & 1 \end{pmatrix} \begin{pmatrix} 1 & 7 & 3 & 2 \\ 0 & -4 & 5 & -1 \\ 0 & 0 & -113/2 & 3/2 \end{pmatrix}$$

At this point we stop because the matrix on the right is in upper triangular form.

3. Find the LU factorization of the coefficient matrix using Dolittle's method and use it to solve the system of equations.

$$\begin{aligned} x + 2y + 3z &= 5 \\ 2x + 3y + 3z &= 6 \\ 3x + 5y + 4z &= 11 \end{aligned}$$

The coefficient matrix is

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 3 \\ 3 & 5 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -3 \\ 0 & 0 & -2 \end{pmatrix}.$$

Then we first solve

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & 1 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} 5 \\ 6 \\ 11 \end{pmatrix}$$

which yields

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} 5 \\ -4 \\ 0 \end{pmatrix}$$

Next we solve

$$\overbrace{\begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & -3 \\ 0 & 0 & -2 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}}^{=(u,v,w)^T} = \begin{pmatrix} 5 \\ -4 \\ 0 \end{pmatrix}$$

which yields

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -3 \\ 4 \\ 0 \end{pmatrix}$$

B.7 Worked Exercises

1. Let M be an $n \times n$ matrix. Then define the adjoint of M,denoted by M^* to be the transpose of the conjugate of M. For example,

$$\begin{pmatrix} 2 & i \\ 1+i & 3 \end{pmatrix}^* = \begin{pmatrix} 2 & 1-i \\ -i & 3 \end{pmatrix}.$$

A matrix, M, is self adjoint if $M^* = M$. Show the eigenvalues of a self adjoint matrix are all real. If the self adjoint matrix has all real entries, it is called symmetric. Show that the eigenvalues and eigenvectors of a symmetric matrix occur in conjugate pairs.

First note that for $\mathbf{x}$ a vector, $\mathbf{x}^*\mathbf{x} = |\mathbf{x}|^2$. This is because

$$\mathbf{x}^*\mathbf{x} = \sum_k \overline{x_k} x_k = \sum_k |x_k|^2 \equiv |\mathbf{x}|^2 .$$

Also note that $(AB)^* = B^*A^*$ because this holds for transposes.This implies that for A an $n \times m$ matrix,

$$\mathbf{x}^* A^* \mathbf{x} = (A\mathbf{x})^* \mathbf{x}$$

Then if $M\mathbf{x} = \lambda\mathbf{x}$

$$\begin{aligned} \overline{\lambda}\mathbf{x}^*\mathbf{x} &= (\lambda\mathbf{x})^*\mathbf{x} = (M\mathbf{x})^*\mathbf{x} = \mathbf{x}^*M^*\mathbf{x} \\ &= \mathbf{x}^*M\mathbf{x} = \mathbf{x}^*\lambda\mathbf{x} = \lambda\mathbf{x}^*\mathbf{x} \end{aligned}$$

and so $\lambda = \overline{\lambda}$ showing that λ must be real.

2. Suppose A is an $n \times n$ matrix consisting entirely of real entries but $a+ib$ is a complex eigenvalue having the eigenvector, $\mathbf{x} + i\mathbf{y}$. Here $\mathbf{x}$ and $\mathbf{y}$ are real vectors. Show that then $a - ib$ is also an eigenvalue with the eigenvector, $\mathbf{x} - i\mathbf{y}$. **Hint:** You should remember that the conjugate of a product of complex numbers equals the product of the conjugates. Here $a + ib$ is a complex number whose conjugate equals $a - ib$.

If A is real then the characteristic equation has all real coefficients. Therefore, letting $p(\lambda)$ be the characteristic polynomial,

$$0 = p(\lambda) = \overline{p(\lambda)} = p\left(\overline{\lambda}\right)$$

showing that $\overline{\lambda}$ is also an eigenvalue.

3. Find the eigenvalues and eigenvectors of the matrix

$$\begin{pmatrix} -10 & -2 & 11 \\ -18 & 6 & -9 \\ 10 & -10 & -2 \end{pmatrix}.$$

Determine whether the matrix is defective.

The matrix has eigenvalues -12 and 18. Of these, -12 is repeated with multiplicity two. Therefore, you need to see whether the eigenspace has dimension two. If it does, then the matrix is non defective. If it does not, then the matrix is defective. The row reduced echelon form for the system you need to solve is

$$\left(\begin{array}{ccc|c} 2 & -2 & 11 & 0 \\ -18 & 18 & -9 & 0 \\ 10 & -10 & 10 & 0 \end{array}\right)$$

and its row reduced echelon form is

$$\left(\begin{array}{ccc|c} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Therefore, the eigenspace is of the form

$$\begin{pmatrix} t \\ t \\ 0 \end{pmatrix}$$

This is only one dimensional and so the matrix is defective.

4. Find the complex eigenvalues and eigenvectors of the matrix $\begin{pmatrix} 1 & 1 & -6 \\ 7 & -5 & -6 \\ -1 & 7 & 2 \end{pmatrix}$.

Determine whether the matrix is defective.

After wading through much affliction you find the eigenvalues are $-6, 2+6i, 2-6i$. Since these are distinct, the matrix cannot be defective. We must find the eigenvectors for these eigenvalues. The augmented matrix for the system of equations which must be solved to find the eigenvectors associated with $2-6i$ is

$$\left(\begin{array}{ccc|c} -1+6i & 1 & -6 & 0 \\ 7 & -7+6i & -6 & 0 \\ -1 & 7 & 6i & 0 \end{array}\right).$$

The row reduced echelon form is

$$\left(\begin{array}{cccc} 1 & 0 & i & 0 \\ 0 & 1 & i & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

and so the eigenvectors are of the form

$$t\begin{pmatrix} -i \\ -i \\ 1 \end{pmatrix}.$$

You can check this as follows

$$\begin{pmatrix} 1 & 1 & -6 \\ 7 & -5 & -6 \\ -1 & 7 & 2 \end{pmatrix}\begin{pmatrix} -i \\ -i \\ 1 \end{pmatrix} = \begin{pmatrix} -6-2i \\ -6-2i \\ 2-6i \end{pmatrix}$$

and

$$(2-6i)\begin{pmatrix} -i \\ -i \\ 1 \end{pmatrix} = \begin{pmatrix} -6-2i \\ -6-2i \\ 2-6i \end{pmatrix}.$$

It follows that the eigenvectors for $\lambda = 2+6i$ are

$$t\begin{pmatrix} i \\ i \\ 1 \end{pmatrix}.$$

This is because A is real. If $A\mathbf{v} = \lambda\mathbf{v}$, then taking the conjugate,

$$A\overline{\mathbf{v}} = \overline{A\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}.$$

It only remains to find the eigenvector for $\lambda = -6$. The augmented matrix to row reduce is

$$\left(\begin{array}{ccc|c} 7 & 1 & -6 & 0 \\ 7 & 1 & -6 & 0 \\ -1 & 7 & 8 & 0 \end{array}\right)$$

The row reduced echelon form is

$$\left(\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

Then an eigenvector is

$$\begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}.$$

5. You own a trailer rental company in a large city and you have four locations, one in the South East, one in the North East, one in the North West, and one in the South West. Denote these locations by SE,NE,NW, and SW respectively. Suppose you observe that in a typical day, .7 of the trailers starting in SE stay in SE, .1 of the trailers in NE go to SE, .1 of the trailers in NW end up in SE, .2 of the trailers in SW end up in SE, .1 of the trailers in SE end up in NE,.7 of the trailers in NE end up in NE,.2 of the trailers in NW end up in NE,.1 of the trailers in SW end up in NE, .2 of the trailers in SE end up in NW, .1 of the trailers in NE end up in NW, .6 of the trailers in NW end up in NW, .2 of the trailers in SW end up in NW, 0 of the trailers in SE end up in SW, .1 of the trailers in NE end up in SW, .1 of the trailers in NW end up in SW, .5 of the trailers in SW end up in SW. You begin with 20 trailers in each location. Approximately how many will you have in each location after a long time? Will any location ever run out of trailers?

   It sometimes helps to write down a table summarizing the given information.

	SE	NE	NW	SW
SE	.7	.1	.1	.2
NE	.1	.7	.2	.1
NW	.2	.1	.6	.2
SW	0	.1	.1	.5

Then the migration matrix is

$$\begin{pmatrix} 7/10 & 1/10 & 1/10 & 1/5 \\ 1/10 & 7/10 & 1/5 & 1/10 \\ 1/5 & 1/10 & 3/5 & 1/5 \\ 0 & 1/10 & 1/10 & 1/2 \end{pmatrix}$$

All we have to do is find the eigenvector (In this case the eigenspace will be one dimensional because some power of the matrix has all positive entries.) corresponding to $\lambda = 1$ which has all the entries add to 20. This will be the long time population. Remember, these processes conserve the sum of the entries. We must row reduce

$$\left(\begin{array}{cccc|c} -3/10 & 1/10 & 1/10 & 1/5 & 0 \\ 1/10 & -3/10 & 1/5 & 1/10 & 0 \\ 1/5 & 1/10 & -2/5 & 1/5 & 0 \\ 0 & 1/10 & 1/10 & -1/2 & 0 \end{array}\right)$$

The row reduced echelon form is

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & -\frac{7}{3} & 0 \\ 0 & 1 & 0 & -\frac{8}{3} & 0 \\ 0 & 0 & 1 & -\frac{7}{3} & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Therefore, the eigenvectors are of the form

$$t \begin{pmatrix} 7 \\ 8 \\ 7 \\ 3 \end{pmatrix}$$

and we simply need to choose t in such a way that the entries add to 20. Thus

$$7t + 8t + 7t + 3t = 20$$

so $t = 4/5$. Then the long time limit equals

$$4/5 \begin{pmatrix} 7 \\ 8 \\ 7 \\ 3 \end{pmatrix} = \begin{pmatrix} 5.6 \\ 6.4 \\ 5.6 \\ 2.4 \end{pmatrix}$$

Thus there will be about 5.6 trailers in SE, 6.4 in NE, 5.6 in NW, and 2.4 in SW. In particular, it appears no location will run out of trailers.

The Fundamental Theorem Of Algebra

The fundamental theorem of algebra states that every non constant polynomial having coefficients in $\mathbb{C}$ has a zero in $\mathbb{C}$. If $\mathbb{C}$ is replaced by $\mathbb{R}$, this is not true because of the example, $x^2+1=0$. This theorem is a very remarkable result and notwithstanding its title, all the best proofs of it depend on either analysis or topology. It was first proved by Gauss in 1797. The proof given here follows Rudin [11]. See also Hardy [6] for a similar proof, more discussion and references. The best proof is found in the theory of complex analysis. Recall De Moivre's theorem from trigonometry which is listed here for convenience.

Theorem C.0.1 *Let $r>0$ be given. Then if n is a positive integer,*

$$[r(\cos t + i\sin t)]^n = r^n(\cos nt + i\sin nt).$$

Recall that this theorem is the basis for proving the following corollary from trigonometry, also listed here for convenience.

Corollary C.0.2 *Let z be a non zero complex number and let k be a positive integer. Then there are always exactly k k^{th} roots of z in $\mathbb{C}$.*

Lemma C.0.3 *Let $a_k \in \mathbb{C}$ for $k=1,\cdots,n$ and let $p(z) \equiv \sum_{k=1}^n a_k z^k$. Then p is continuous.*

Proof:

$$|az^n - aw^n| \le |a||z-w|\left|z^{n-1}+z^{n-2}w+\cdots+w^{n-1}\right|.$$

Then for $|z-w|<1$, the triangle inequality implies $|w|<1+|z|$ and so if $|z-w|<1$,

$$|az^n - aw^n| \le |a||z-w|\,n(1+|z|)^n.$$

If $\varepsilon>0$ is given, let

$$\delta < \min\left(1, \frac{\varepsilon}{|a|\,n(1+|z|)^n}\right).$$

It follows from the above inequality that for $|z-w|<\delta$, $|az^n-aw^n|<\varepsilon$. The function of the lemma is just the sum of functions of this sort and so it follows that it is also continuous.

Theorem C.0.4 *(Fundamental theorem of Algebra) Let $p(z)$ be a nonconstant polynomial. Then there exists $z\in\mathbb{C}$ such that $p(z)=0$.*

Proof: Suppose not. Then

$$p(z) = \sum_{k=0}^{n} a_k z^k$$

where $a_n \neq 0,\ n > 0$. Then

$$|p(z)| \geq |a_n| |z|^n - \sum_{k=0}^{n-1} |a_k| |z|^k$$

and so

$$\lim_{|z|\to\infty} |p(z)| = \infty. \tag{3.1}$$

Now let

$$\lambda \equiv \inf \{|p(z)| : z \in \mathbb{C}\}.$$

By 3.1, there exists an $R > 0$ such that if $|z| > R$, it follows that $|p(z)| > \lambda + 1$. Therefore,

$$\lambda \equiv \inf \{|p(z)| : z \in \mathbb{C}\} = \inf \{|p(z)| : |z| \leq R\}.$$

The set $\{z : |z| \leq R\}$ is a closed and bounded set and so this infimum is achieved at some point w with $|w| \leq R$. A contradiction is obtained if $|p(w)| = 0$ so assume $|p(w)| > 0$. Then consider

$$q(z) \equiv \frac{p(z+w)}{p(w)}.$$

It follows $q(z)$ is of the form

$$q(z) = 1 + c_k z^k + \cdots + c_n z^n$$

where $c_k \neq 0$, because $q(0) = 1$. It is also true that $|q(z)| \geq 1$ by the assumption that $|p(w)|$ is the smallest value of $|p(z)|$. Now let $\theta \in \mathbb{C}$ be a complex number with $|\theta| = 1$ and

$$\theta c_k w^k = -|w|^k |c_k|.$$

If

$$w \neq 0, \theta = \frac{-\left|w^k\right| |c_k|}{w^k c_k}$$

and if $w = 0$, $\theta = 1$ will work. Now let $\eta^k = \theta$ and let t be a small positive number.

$$q(t\eta w) \equiv 1 - t^k |w|^k |c_k| + \cdots + c_n t^n (\eta w)^n$$

which is of the form

$$1 - t^k |w|^k |c_k| + t^k (g(t,w))$$

where $\lim_{t\to 0} g(t,w) = 0$. Letting t be small enough,

$$|g(t,w)| < |w|^k |c_k| /2$$

and so for such t,

$$|q(t\eta w)| < 1 - t^k |w|^k |c_k| + t^k |w|^k |c_k| /2 < 1,$$

a contradiction to $|q(z)| \geq 1$. This proves the theorem.

Bibliography

[1] **Apostol T.** *Calculus Volume II Second edition,* Wiley *1969.*

[2] **Baker, Roger**, *Linear Algebra*, Rinton Press 2001.

[3] **Davis H. and Snider A.,** *Vector Analysis* Wm. C. Brown 1995.

[4] **Edwards C.H.** *Advanced Calculus of several Variables,* Dover 1994.

[5] **Gurtin M.** *An introduction to continuum mechanics,* Academic press 1981.

[6] **Hardy G.** *A Course Of Pure Mathematics, Tenth edition,* Cambridge University Press 1992.

[7] **Horn R. and Johnson C.** *matrix Analysis,* Cambridge University Press, 1985.

[8] **Karlin S. and Taylor H.** *A First Course in Stochastic Processes,* Academic Press, 1975.

[9] **Kuttler K.***Linear Algebra*On web page. Linear Algebra

[10] **Nobel B. and Daniel J.** *Applied Linear Algebra,* Prentice Hall, 1977.

[11] **Rudin W.** *Principles of Mathematical Analysis*, McGraw Hill, 1976.

[12] **Salas S. and Hille E.,** *Calculus One and Several Variables,* Wiley 1990.

[13] **Strang Gilbert**, *Linear Algebra and its Applications,* Harcourt Brace Jovanovich 1980.

[14] **Yosida K.,** *Functional Analysis, Springer Verlag, 1978.*

Index